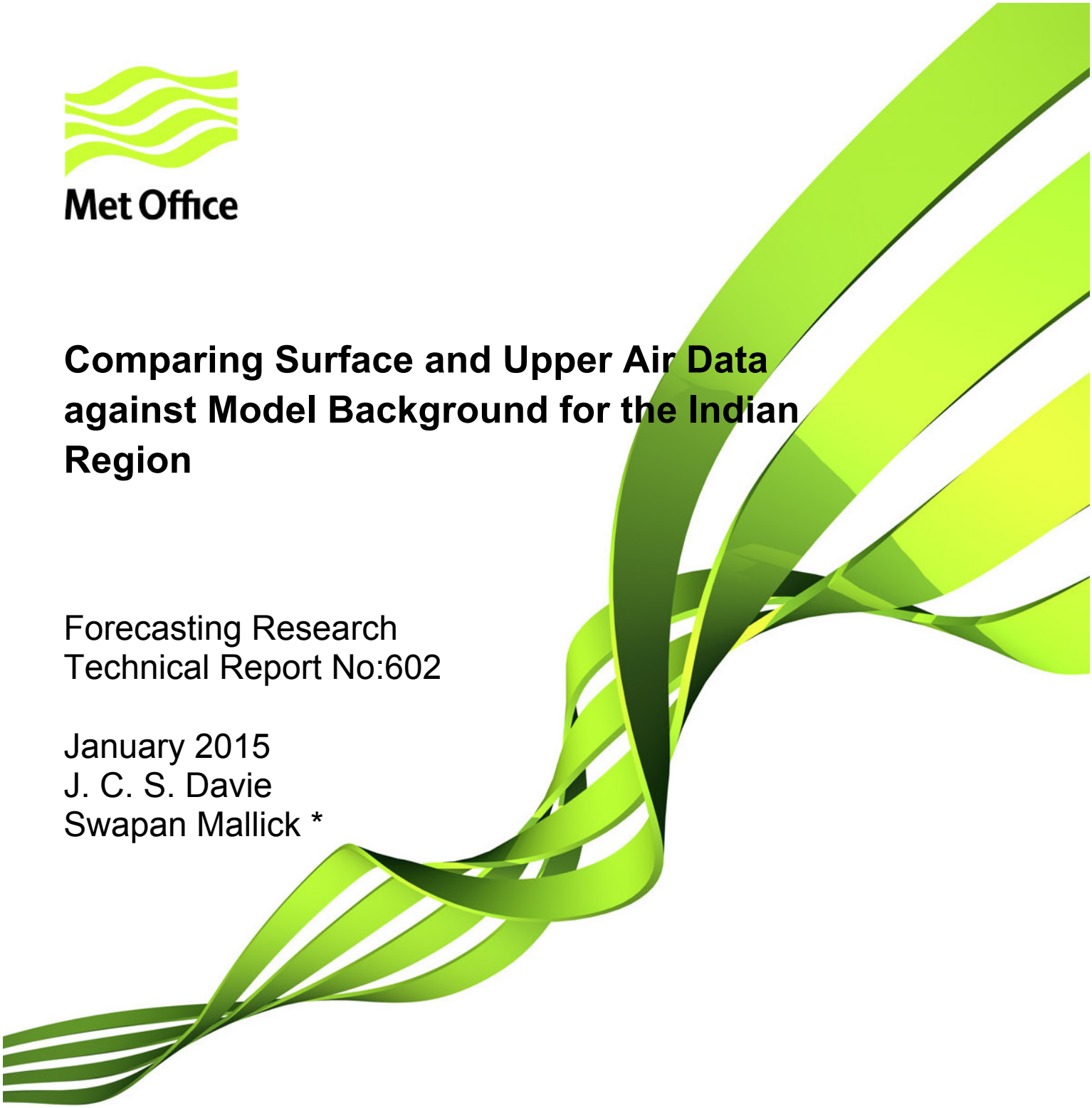

Comparing Surface and Upper Air Data against Model Background for the Indian Region

Forecasting Research
Technical Report No:602

January 2015

J. C. S. Davie
Swapan Mallick *

* National Centre for Medium Range Weather Forecasting (NCMRWF), India

ACKNOWLEDGEMENTS

Colin Parrett produced the plots shown in Figures 38 to 42.

Discussion with Richard Renshaw, Peter Jermey, Sana Mahmood, Colin Parrett and Bruce Ingleby has helped produce this report.

This work was carried out as part of the IMDAA (Indian Monsoon Data Assimilation and Analysis) project, funded by the National Monsoon Mission of India.

Second author acknowledges the guidance and help received from NCMRWF Scientists, specially John P. George and Munmun Das Gupta, while carrying out the work reported in the study, at NCMRWF.

© Crown copyright 2014 Met Office

INTRODUCTION

The purpose of the report is to compare surface and upper air data against model background for the Indian region.

The premise is that there are many observations which are not assimilated over India. In particular, winds are not assimilated in the Tropics. Ingleby (2014) shows that there are biases in daytime wind speed observations from SYNOP reports, which may be a combination of model and observation errors.

This report aims to investigate O-B statistics to explore these issues.

The first section of the report will outline the usage of Indian observations, based on top level decisions used in station list updates. An example of numbers of stations and records saved in Observational DataBase (ODB) for observations in the region are also given for the year May 2013 to April 2014.

The second section of the report will briefly discuss the methods used in the third and fourth sections, which will investigate O-B statistics for surface and upper air data respectively. Section 3 is split into two parts; the first considering the region as a whole, and the second considering individual stations within the region. Section 4 shows examples of both good and poor quality stations based on their O-B profiles for a number of variables.

Finally, the fifth section of the report will summarise and conclude from the preceding sections. It also describes possible ways to extend the analysis done in this report comparing surface and upper air data with model background.

© Crown copyright 2014 Met Office

DATA USAGE

Top level decisions used for station list updating are the following:

Surface data

- Use winds over land from SYNOP and METAR
- Don't use winds from MOBSYNOP
- Reject TEST ship reports
- Reject corrupt ship reports with BBXX in callsign
- Reject ships if incorrectly positioned at the Equator
- Reject SYNOP and METAR winds from 30S to 30N
- Reject marine reports over land
- Reject MOBSYNOP from 65E to 90E, 5N to 35N

This means that temperature, relative humidity and pressure from SYNOP, METAR, SHIP and BUOY are used over India, but not from MOBSYNOP between 65E to 90E and 5N to 35N. Winds from SYNOP and METAR are rejected from 30S to 30N.

Upper Air data

- Thin profilers to 1 report per station per hour
- Thin PILOT reports from each station to 1 per 2 hour 2 minutes
- Reject obstypes 50500 and 50400 and reports of temperature, RH and X from obstypes beginning with 502, TEMP ship and PILOT ship over land, 'eyewall' dropsondes (there is a list), TEST reports, reports at 0N, 0E, sonde reports from non-radiosonde systems

For TEMPs, there are 8 stations with their temperature and relative humidity ascents currently rejected, and 2 of these stations also have their wind ascents rejected (C. Parrett, personal communication).

These decisions are used along with decisions made based on the O-B statistics to create the station lists containing stations to reject.

© Crown copyright 2014 Met Office

Numbers of surface stations/records over India

From ODB for May 2013 to April 2014, the number of stations with records and the number of records for each variable was:

	SYNOP	MOBSYN	SHIP	BUOY	METAR
Number of stations	505	695	736	87	84
Number of records	880275	8525	27177	115373	625530

The number of records with missing observation values in the ODB over India was:

	SYNOP	MOBSYN	SHIP	BUOY	METAR
2m temperature	602	1510	575	68779	2729
Surface pressure	26685	8525	209	35622	24775
Wind speed (UV)	2808	8525	530	76173	51357
Relative humidity	2018	8525	3007	108062	3225

The number of active records (datum_status.active = 1) over India was:

	SYNOP	MOBSYN	SHIP	BUOY	METAR
2m temperature	763886	0	25992	45518	567015
Surface pressure	798745	0	25218	78892	598981
Wind speed (UV)	443929	0	24746	37332	501465
Relative humidity	760539	0	20009	7302	557611

The number of rejected records (datum_status.rejected = 1) was:

	SYNOP	MOBSYN	SHIP	BUOY	METAR
2m temperature	116389	8525	1185	69855	58515
Surface pressure	81530	8525	1959	36481	26549
Wind speed (UV)	436346	8525	2431	78041	124065
Relative humidity	119736	8525	7168	108071	67919

The number of blacklisted records (datum_status.blacklisted = 1) was:

	SYNOP	MOBSYN	SHIP	BUOY	METAR
2m temperature	113844	8525	356	67870	472144
Surface pressure	47047	223	739	5951	444421
Wind speed (UV)	734434	8525	1286	69312	549289
Relative humidity	110099	8525	4132	67039	470324

© Crown copyright 2014 Met Office

METHODS

Surface

The analysis for surface data included in this report uses archived ODB (Fouilloux, 2010) from the operational global model as the input, and includes the fields of lat, lon, date, time, antime, andate, statid, stalt, obsvalue, fg_depar, varno, datum_status.active, datum_status.passive, datum_status.blacklisted, datum_status.rejected, ops_obstype, an_depar, report_pge and ob_practice.

The variables considered are 2m temperature, 10m wind speed, surface pressure and 2m relative humidity. The 10m wind speed has been derived from the 10m u- and v-components.

The surface observation types have been grouped as SYNOP (10100, 10101, 10102), Mobile SYNOP (10800), Metar (11101, 11102), Ship (10200, 10201, 10202) and Buoy (10210, 10211, 10212, 10213, 10300, 10310, 10311, 10312, 10313).

The region used is 60E – 100E and 5N – 40N.

Observation records where the observed value or the departure value in the ODB is set to missing data values are ignored and not included in the analysis.

Upper Air

The variables considered from TEMP reports are wind speed, temperature, wind vector and relative humidity. We have not considered PILOTs or wind profilers in this report.

© Crown copyright 2014 Met Office

RESULTS
Surface Data Region O and B comparison

2m Temperature

The mean and median O-B over the region for each month remained fairly constant over the year May 2013 to April 2014, at approximately 1K. The spread in O-B values slightly reduced between July and September 2013, with values ranging from ~-2.5K to ~5K (Figure 1). Time series for buoys show that the O-B values are centred closer to 0, with less spread, but there were fewer stations. For ships, O-B values were centred at 0, with less spread during winter months, and a slight reduction in O-B value compared to the rest of the year (figure not shown).

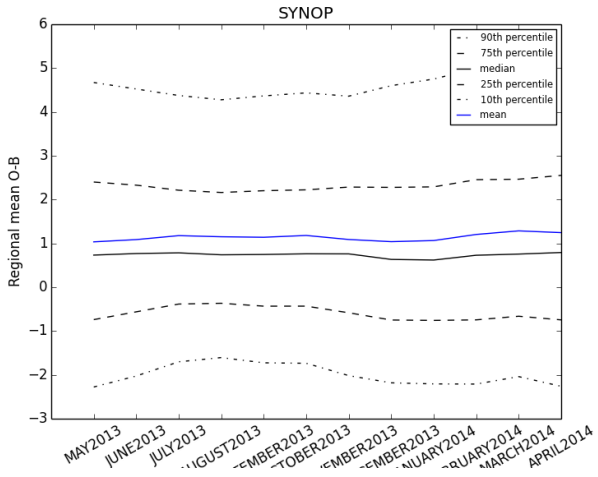

Figure 1: Time series of monthly 90th percentile, 75th percentile, median, 25th percentile, 10th percentile and mean of 2m temperature (degrees K) O-B values for the region 60E-100E and 5N-40N

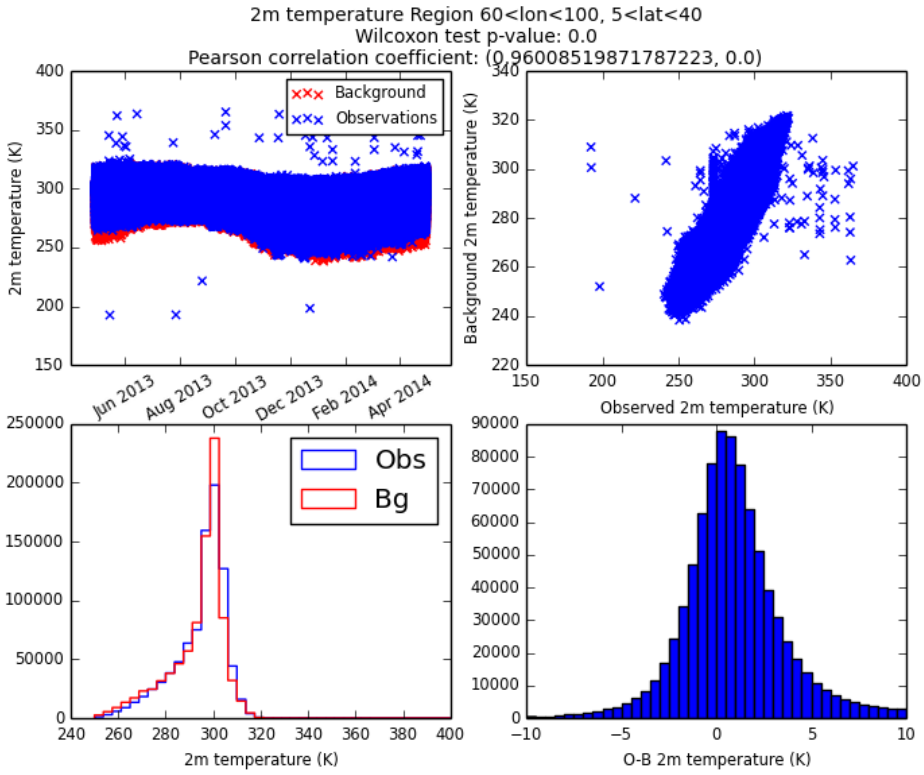

Figure 2: Scatter plot of observed and background 2m temperature values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

Figure 2 shows that observed values are generally slightly higher than background values of 2m temperature, but that they have strong positive correlation, with a Pearson correlation coefficient value of 0.96. The very small Wilcoxon signed rank test (Scipy.org) p-value however suggests that the observed values and background values are not from the same distribution, as the alternative hypothesis that the two sets of values are not from the same distribution would be accepted in preference to the null hypothesis that they are from the same distribution. The histograms also show a higher frequency of lower background values and of higher observed values, and a fairly symmetrical distribution of O-B values centred just above 0K. Buoy and ship (not shown) O-B distributions differ to SYNOP, with buoys exhibiting a weaker positive correlation between O and B values, more observed values at the upper and lower ends of the range, and a longer negative tail to the O-B distribution, while ships show a longer positive tail to the O-B distribution.

The broad spatial pattern of values is similar for monthly mean observed and background values (not shown). Temporal mean O-B values are mainly between -1K and 1K, with a few stations with more extreme values, but not obviously more negative or more positive values (Figure 3). The north of the region has many high O-B values over the Himalayas which may be biasing regional mean, median etc. to suggest observed values are warmer than the background.

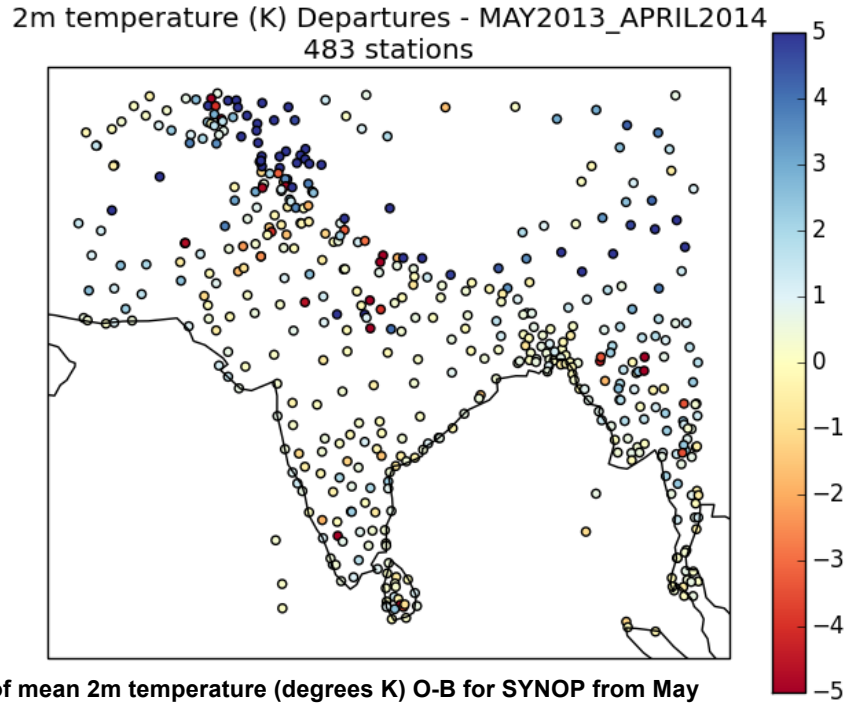

Figure 3: Map of mean 2m temperature (degrees K) O-B for SYNOP from May 2013 to April 2014 for the region 60E-100E, 5N-40N

Not shown are equivalents of Figures 2 and 3 for each month during the period. These show that the correlation between observed and background values is more strongly positive during the winter than the summer, with Pearson correlation coefficients of ~0.95 compared to ~0.9. The O-B values are centred just above 0 and the maps show no obvious changes spatially throughout the year considered.

© Crown copyright 2014 Met Office

Surface Pressure

Figure 4 shows a slight downward shift in the distribution of O-B values between May 2013 and April 2014, centred approximately at 0 as the mean and median and values ranging from -1.5hPa to 1.5hPa. There are higher mean O-B for ships and buoys. Buoys, similarly to SYNOP, have a slight shift to lower O-B values over the time period considered and a reduction in spread (not shown). The mean O-B for ships remains fairly constant, although there is reduced spread during summer months.

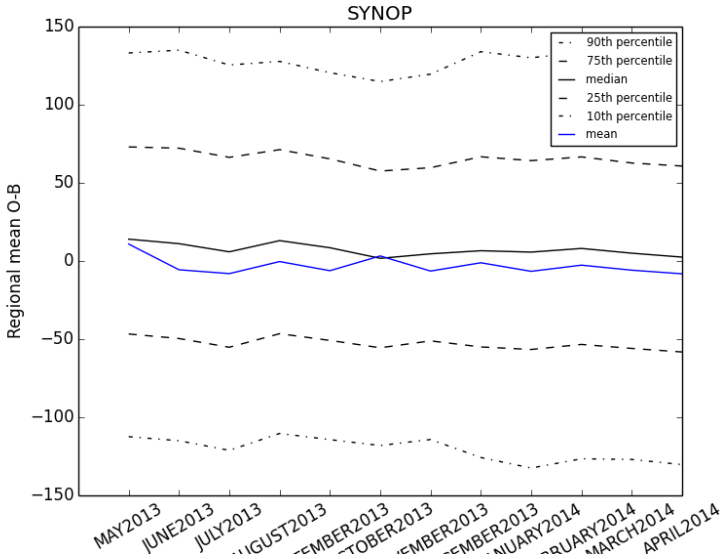

Figure 4: Time series of monthly 90th percentile, 75th percentile, median, 25th percentile, 10th percentile and mean of surface pressure (Pa) O-B values for the region 60E-100E and 5N-40N

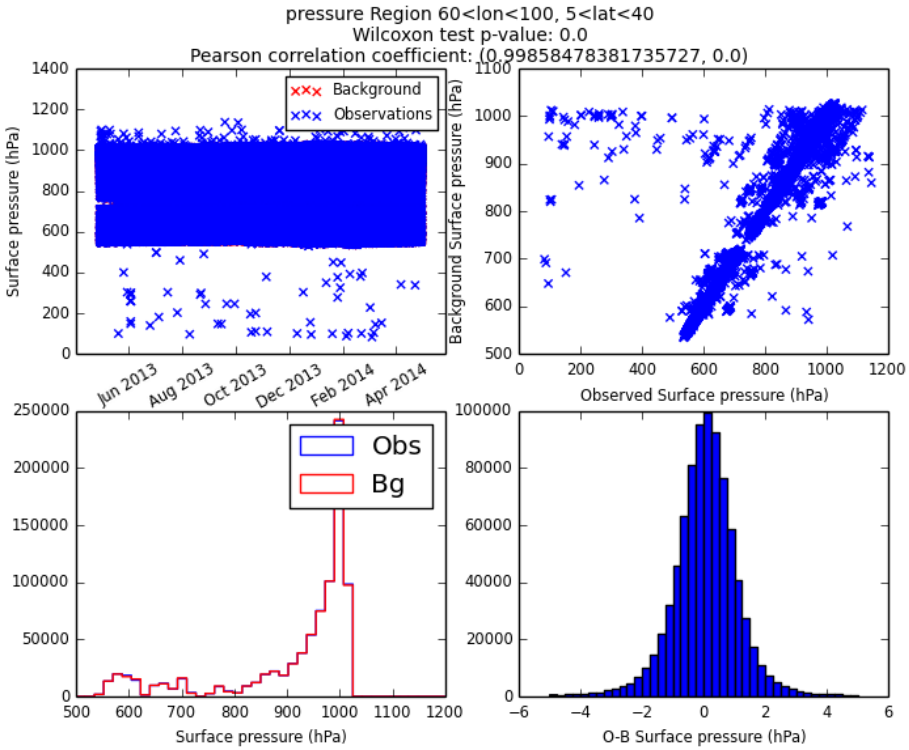

Figure 5: Scatter plot of observed and background surface pressure values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

The left column of plots in Figure 5 show close similarity between observed and background values of surface pressure, despite which, the Wilcoxon signed rank test suggests that they are not likely to be from the same distribution. There is a very strong positive correlation, 0.999 between the two sets of values and the histogram of O-B values is bell-shaped and symmetrical about 0. There are also strong positive correlations exhibited between O and B values for ships and buoys and symmetrical distributions of O-B (not shown).

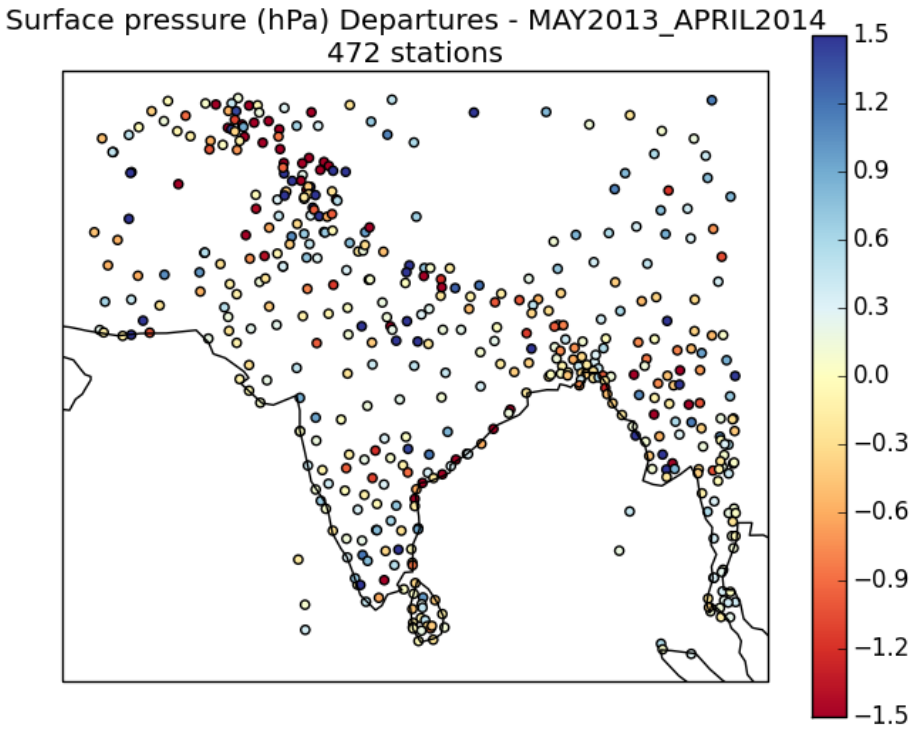

Figure 6: Map of mean surface pressure (hPa) O-B for SYNOP from May 2013 to April 2014 for the region 60E-100E, 5N-40N

The observed and background values have similar spatial patterns, with a few low observed values on the NE coastline and a couple of high observed values in central Northern India (not shown). The temporal mean O-B values are very mixed, with no clear spatial pattern (Figure 6).

There are no clear patterns spatially of the mean O-B pressure values varying each month, and the histograms are also similar over the year (not shown).

© Crown copyright 2014 Met Office

10m Wind Speed

The majority (~80%) of the O-B values are negative, so the observed values are generally lower than the corresponding background, with 50% of the O-B values below -1.5m/s (Figure 7). There appears to be a slight reduction in magnitude of O-B values throughout the year with spread also reducing from July 2013, mainly by a reduction in highly negative values. The distributions of O-B values for ships and buoys differ to SYNOP as O-B values from buoy reports have a symmetrical distribution centred at 0, and the distribution for ships is centred just above 0.

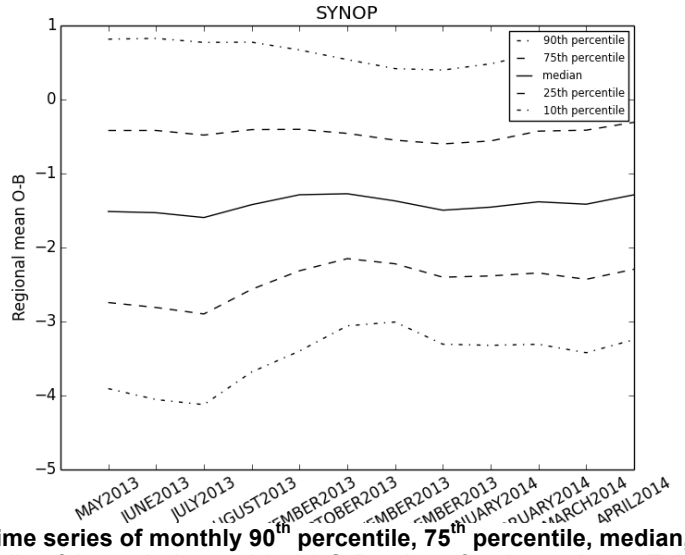

Figure 7: Time series of monthly 90th percentile, 75th percentile, median, 25th percentile and 10th percentile of 10m wind speed (m/s) O-B values for the region 60E-100E and 5N-40N

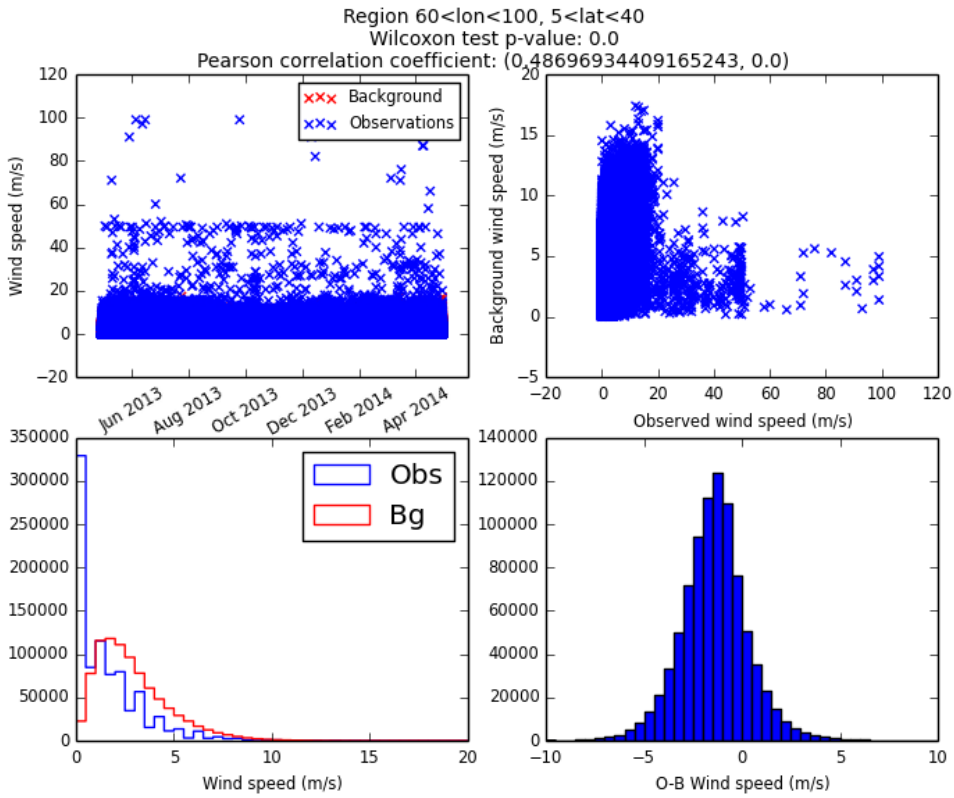

Figure 8: Scatter plot of observed and background 10m wind speed values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

There is a moderately positive correlation of 0.487 between the observed and background values, and there is a much higher frequency of low observed wind speeds than of low background values, although there are similar maximum values (Figure 8). The background has a smoother distribution, which may be because the wind speeds are recorded with less accuracy. The O-B values form a fairly symmetrical distribution, centred below 0. This may be unduly influenced by the large number of calm winds recorded, so these results may change if they were left out of the analysis. There is also moderately positive correlation between observed and background values for ships, but a strong positive correlation for buoys which have a symmetrical O-B distribution centred about 0 (not shown).

Figure 9 shows that temporal mean O-B values are predominantly negative across the region, with a large proportion of stations having negative mean O-B. However, there are a few coastal stations with positive values.

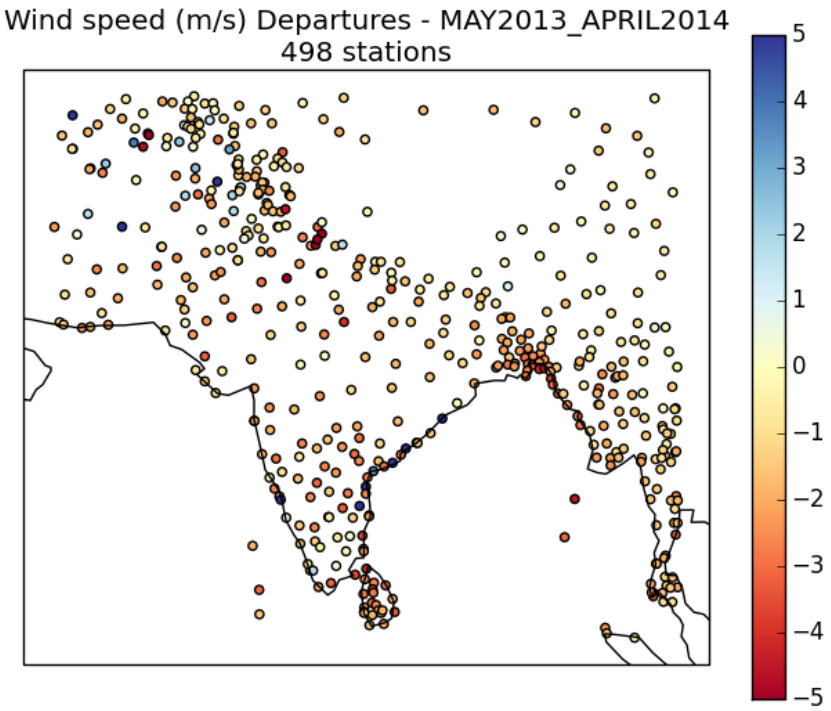

Figure 9: Map of mean wind speed (m/s) O-B for SYNOP from May 2013 to April 2014 for the region 60E-100E, 5N-40N

The correlation coefficients are similar all year. However, the distribution of O-B appears negatively biased during the summer, and the maps also show more stations with highly negative monthly mean O-B values during the summer months.

© Crown copyright 2014 Met Office

The high negative biases in the surface wind observations over Indian region, has been also analysed with respect to background (B) from UM run (NCUM) made at National Centre for Medium Range Weather Forecasting (NCMRWF) (Rajagopal et al., 2012) for August' 2013. The number of surface observations over India received and saved at NCMRWF observation data bank, and used for this study are almost similar to that at Met Office described above(first section), and hence not listed here for brevity. An attempt has been made to compare surface wind bias for SYNOP observations and automatic weather station observations (SYNOPMOB) over Indian region. Figure 9.1 shows the monthly average for day and night time of O-B for SYNOP and AWS observations for August 2013.

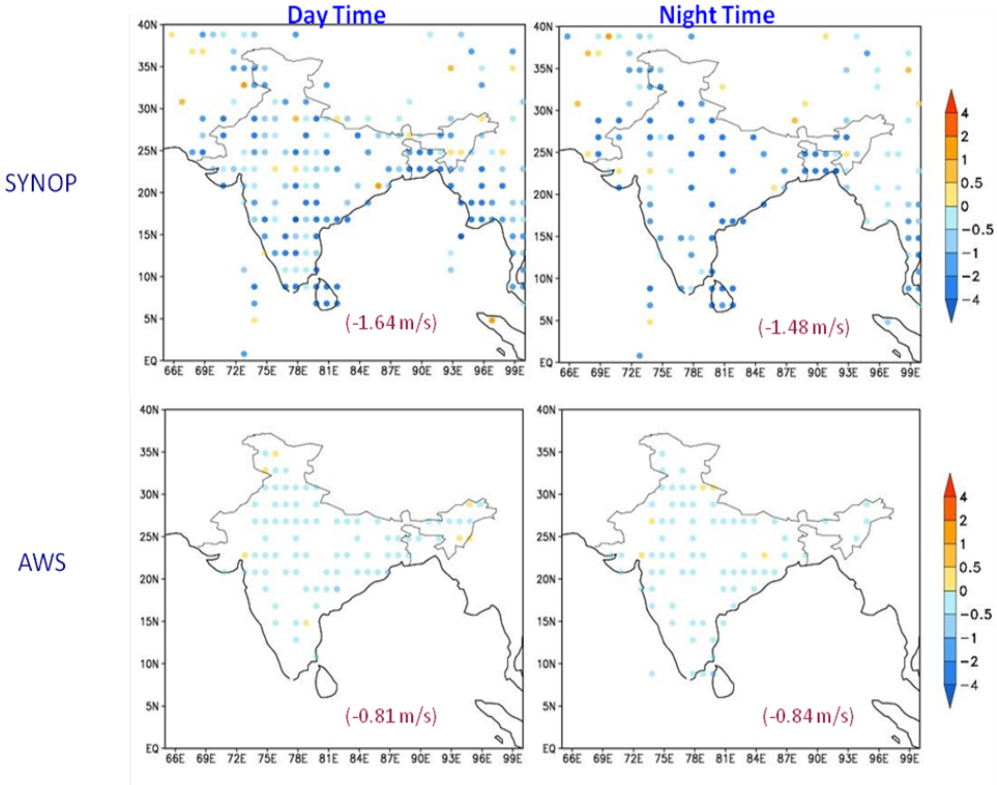

Figure 9.1: Monthly average day time and night time O-B for observed surface (SYNOP top two panels, AWS bottom two panels) and NCUM background wind speed for August 2013 over India. The number in each panel represents average O-B for August 2013.

As seen from the plot, for both SYNOP and AWS, O-B is predominantly negative over Indian region. However, for SYNOP the value of O-B is almost double to that of AWS. Not much difference is seen between O-B values computed for day time and night time for both SYNOP and AWS observations over India, which is in the line of the results found in the study carried by Ingleby, 2014.

Kernel density estimates (Wilks, 2006) of observed surface wind speed along with NCUM background of 10 m wind speed for the month of August 2013 for Indian SYNOP & AWS observations are shown in figure 9.2. It is seen that the density low wind (0-1 m s^{-1}) is higher in observation than the model background for both SYNOP and AWS. However, the density of calm wind (speed~0 m s^{-1}) is more in

© Crown copyright 2014 Met Office

SYNOP observations compared to that in AWS observations. In the medium wind speed range (2.5 to 5.5 m s $^{-1}$) the density of observed wind is substantially less than that of background over Indian region for both SYNOP and AWS.

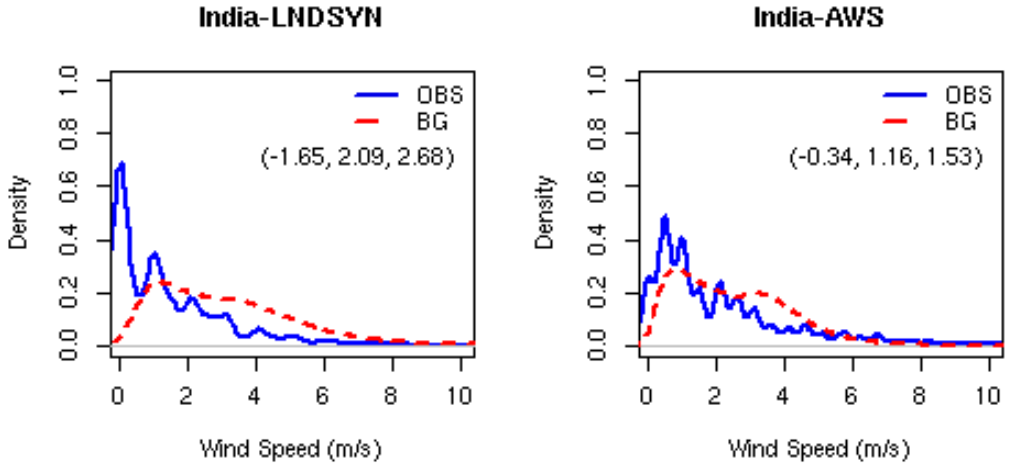

Figure 9.2: Kernel density estimates for observed surface and NCUM background wind speed for August 2013 for Indian SYNOP and AWS. The numbers in each panel represent respectively, average O-B, average absolute error and root mean square error for August 2013.

2m Relative Humidity

The monthly O-B relative humidity values are spread between approximately -20% and 15% with mean and median both just below 0 and there is a reduction in spread from May until September 2013 (Figure 10).

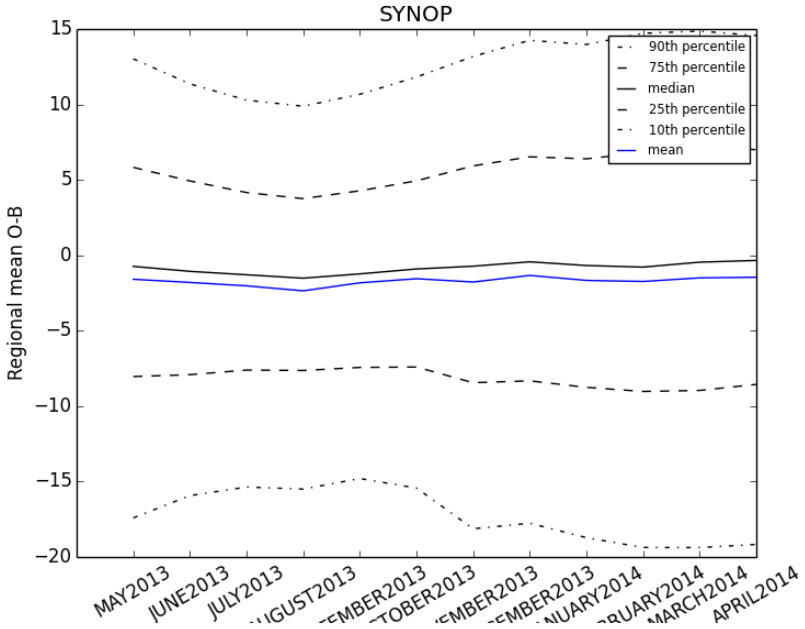

Figure 10: Time series of monthly 90th percentile, 75th percentile, median, 25th percentile and 10th percentile of 2m relative humidity (%) O-B values for the region 60E-100E and 5N-40N

© Crown copyright 2014 Met Office

The distribution of O-B values for buoy reports increased from January (mean ≈ 1.5) to April 2014 (mean ≈ 8) and for ships the mean was below 0 between June and September 2013, but increased to above 0 afterwards, with a symmetrical distribution (not shown).

Figure 11 shows that the observed and background values of 2m relative humidity are moderately positively correlated. The plots in the left column show that more of the observed values than background values are between 30% and 70% and there are more background values at the higher end of the range than observed values. The distribution of O-B values is fairly symmetrical, peaking close to 0%, although there is a longer tail to the histogram in the negative direction suggesting that there are more records where the observed value is lower than the background value. Buoys exhibit a strong positive correlation between observed and background relative humidity values, although there are more high observed values and more low background values (not shown). Ships also have more high observed values, but there is only weak positive correlation (not shown).

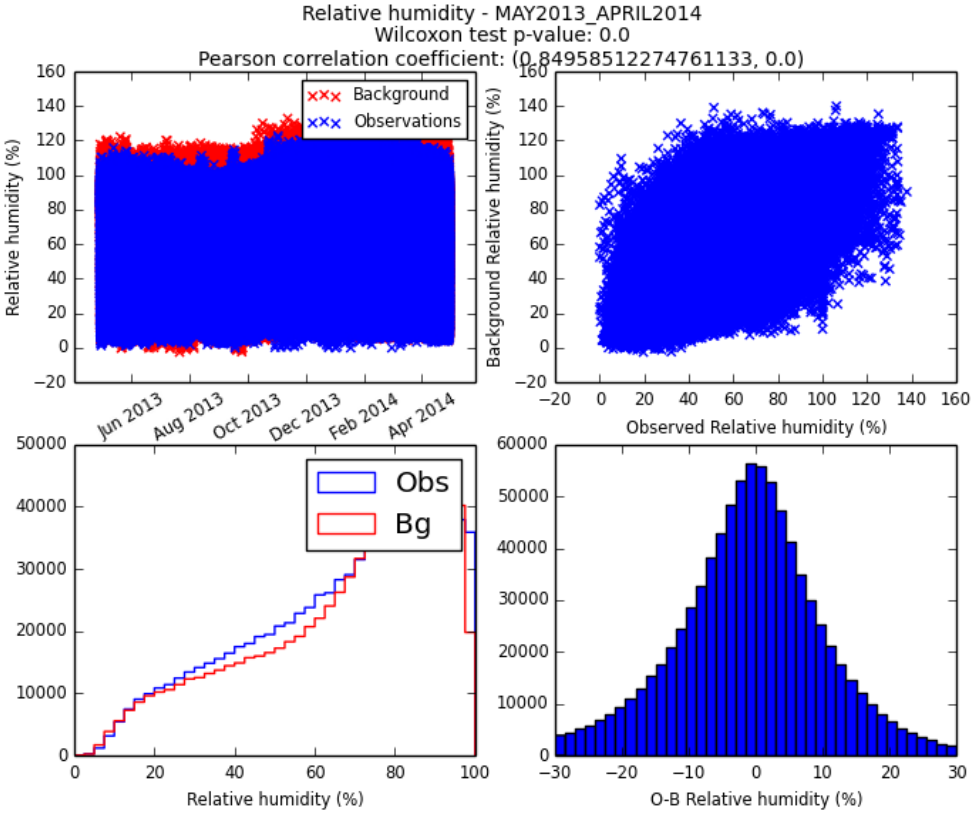

Figure 11: Scatter plot of observed and background 2m relative humidity values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

There is a mixture of stations with positive and negative mean O-B values (Figure 12). The majority of the stations with highly negative values are over the Himalayas. This means that the distribution of O-B values for the region as shown in Figures 10 and 11 could be quite different if the region considered was limited to a smaller region. There appear to be more stations with positive O-B monthly means during winter and more with negative O-B monthly means during summer (not shown).

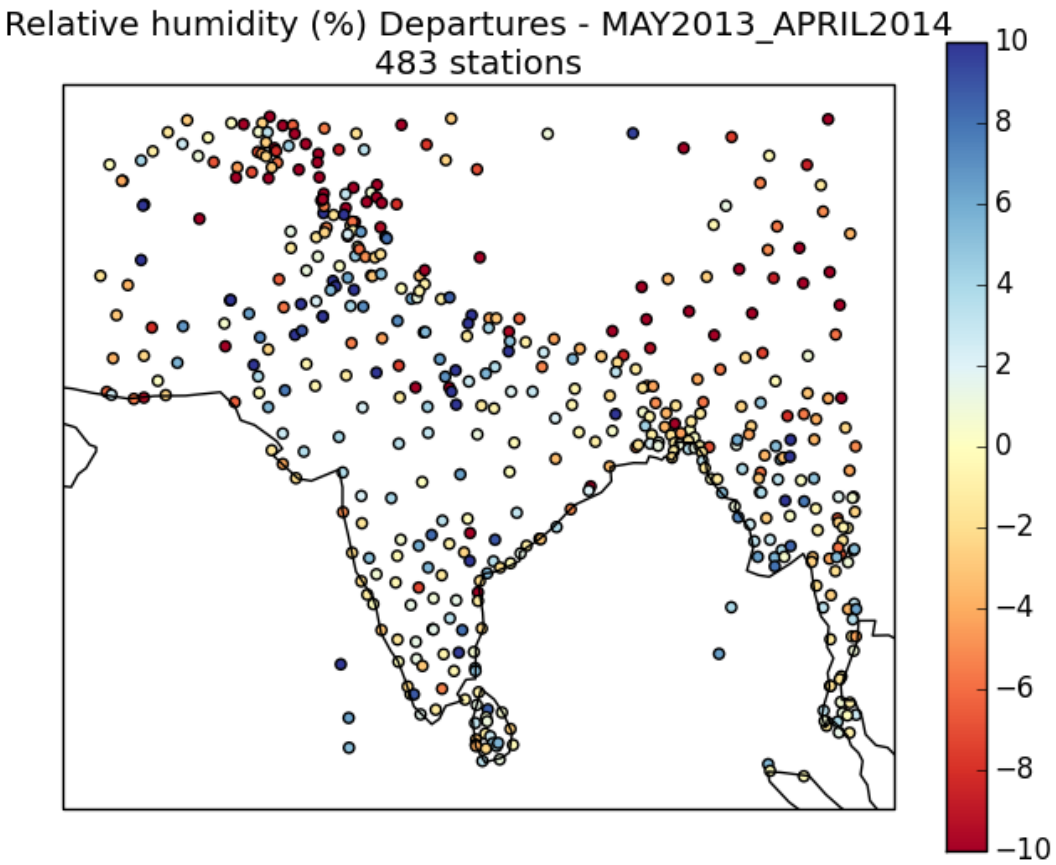

Figure 12: Map of mean 2m relative humidity (%) O-B for SYNOP from May 2013 to April 2014 for the region 60E-100E, 5N-40N

© Crown copyright 2014 Met Office

Surface Station O and B comparison

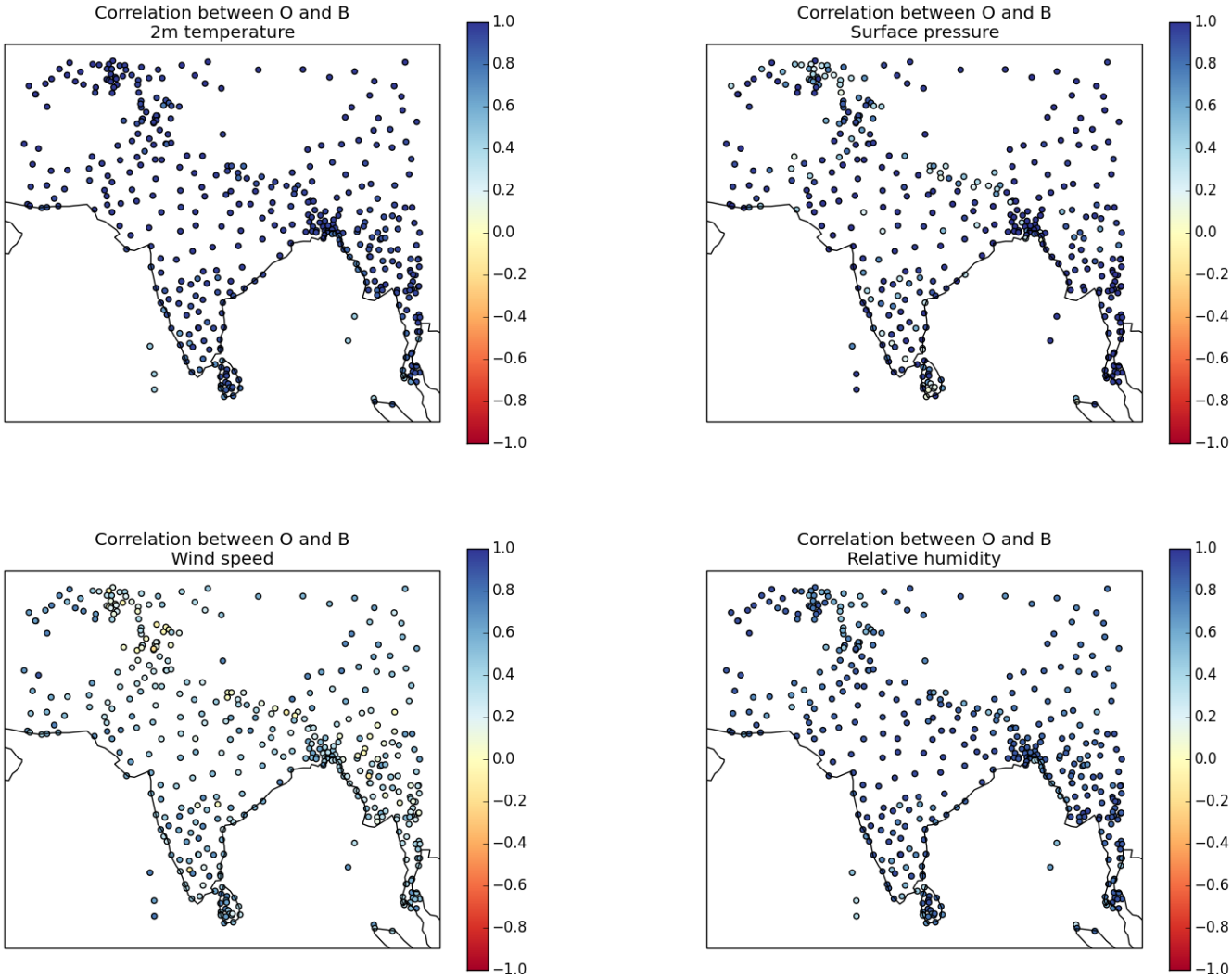

Figure 13: Pearson correlation coefficient values between observed and background values from May 2013 until April 2014 – 2m temperature (upper left), surface pressure (upper right), wind speed (lower left) and 2m relative humidity (lower right)

Figure 13 shows the Pearson correlation coefficient for observed compared with background values, at each station with more than 20 SYNOP reports per month, from May 2013 until April 2014. Higher values imply a stronger relationship between observed and background values, with 1 being a perfect positive linear relationship. For both 2m temperature and 2m relative humidity, there are strong positive correlations, although island stations are weaker. The majority of stations have strong positive correlations between observed and background surface pressure values. A few stations exhibit weaker correlation, but these are mostly in mountainous areas in the north of the region. There tend to be weaker correlations for 10m wind speed, which have moderate positive correlation at approximately half of the stations, although there are many which are uncorrelated, particularly in mountainous areas. Most stations situated in the Indian peninsula have strong positive correlations.

© Crown copyright 2014 Met Office

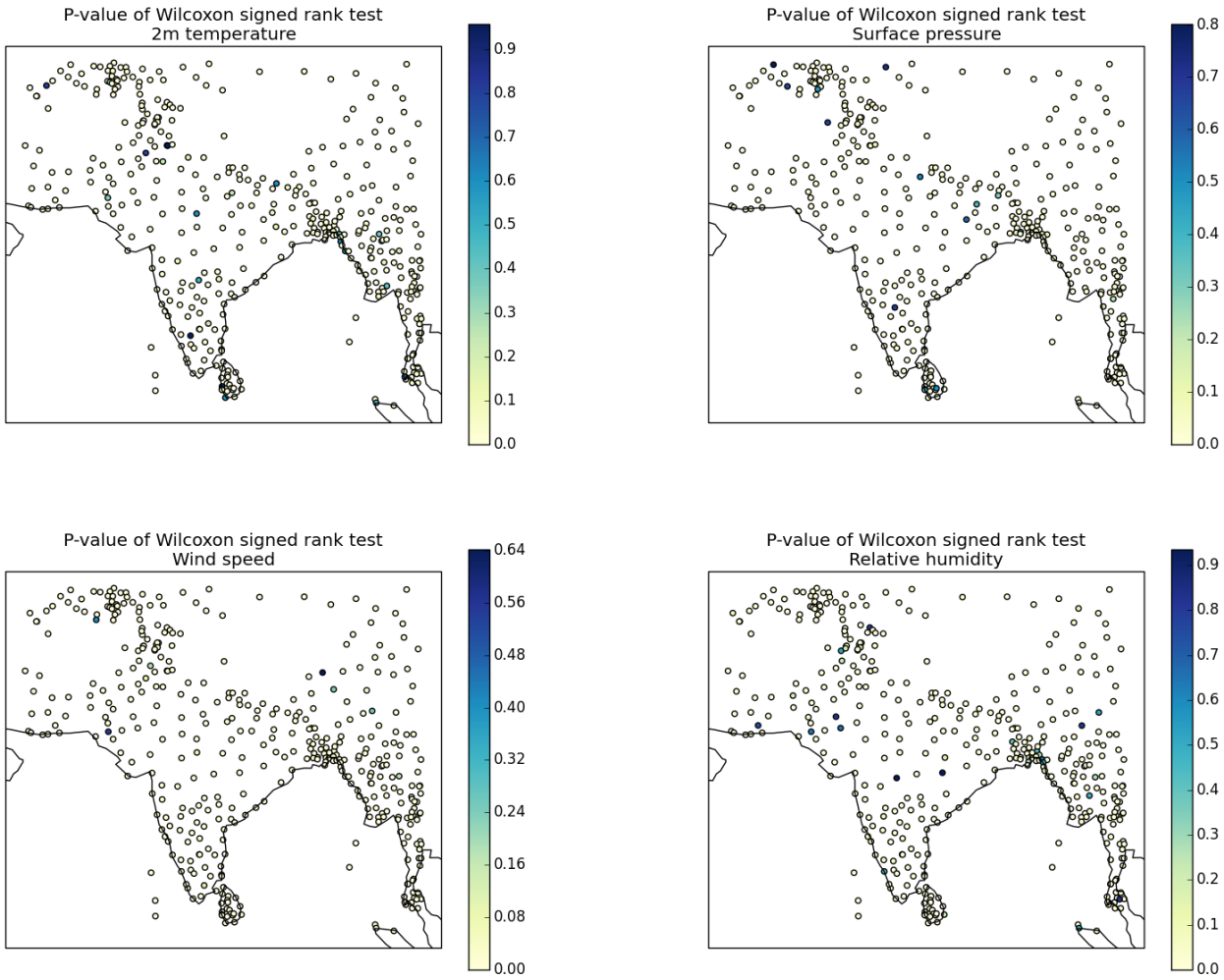

Figure 14: P-values from Wilcoxon signed rank test between observed and background values from May 2013 until April 2014 – 2m temperature (upper left), surface pressure (upper right), wind speed (lower left) and 2m relative humidity (lower right)

For all four variables considered in Figure 14, there were very few stations where it was likely that the observed values and background values from May 2013 until April 2014 came from the same distribution. This is shown using the p-values from a Wilcoxon signed rank test. If the values are low, then the alternative hypothesis that the two sets of values are not from the same distribution would be accepted in preference to the null hypothesis that they are likely to be from the same distribution. There are many reasons which the values may not be from the same distribution, such as low correlation, bias and differences in range of values. The variable with the fewest stations that have p-values greater than 0.05 is wind speed; all stations on the Indian peninsula have very low p-values so observed and background values are unlikely to be from the same distribution. There are a few stations in the north of the Indian peninsula which for relative humidity have p-values greater than 0.3. Figure 14 considers values for a whole year, but there may be some months which have more similar observed and background values than others as the model may produce a better background during particular times of year or particular weather conditions prevalent during different seasons may be observed better than others.

© Crown copyright 2014 Met Office

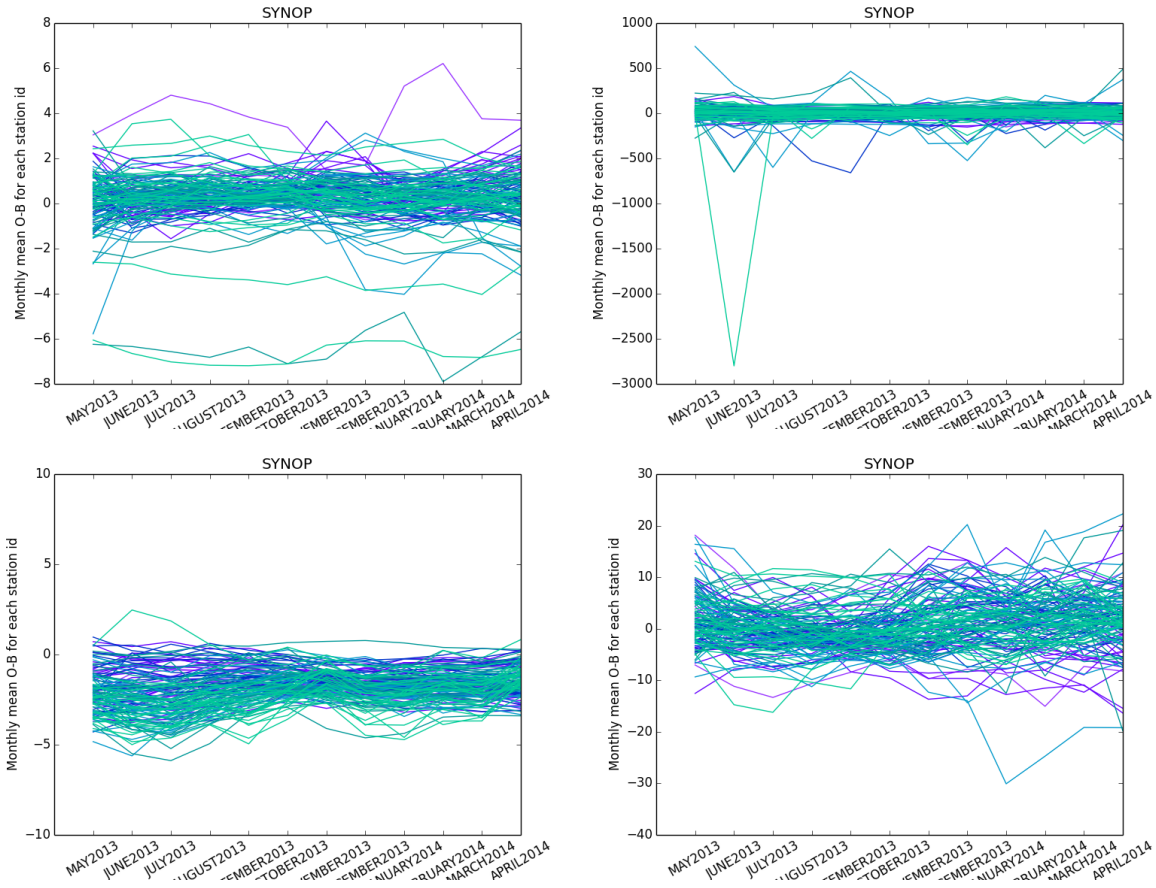

Figure 15: Monthly mean O-B for each individual SYNOP station with different colours representing different latitudinal bands of the region. 2m temperature (upper left), surface pressure (upper right), 10m wind speed (lower left) and 2m relative humidity (lower right). Green lines are low latitudes fading through to pink lines for the high latitudes.

There appears to be no obvious pattern with latitude for both surface pressure and 2m relative humidity station monthly mean O-B (Figure 15). This is also the case with station altitude (Figure 16). There is reduced spread in stations' monthly mean O-B 2m relative humidity values during summer months (Figure 15). For 2m temperature, the stations with highly negative monthly mean O-B values seem to be mainly at low latitudes in the southern part of the region, and most extreme values appear to be stations at high altitudes. The most negative monthly mean O-B values for 10m wind speed during the summer months (May to September 2013) are at lower latitudes, and there is also more spread between stations' values (Figure 15). The stations exhibiting the most negative monthly mean 10m wind speed O-B values during the winter months were at low altitudes (Figure 16).

© Crown copyright 2014 Met Office

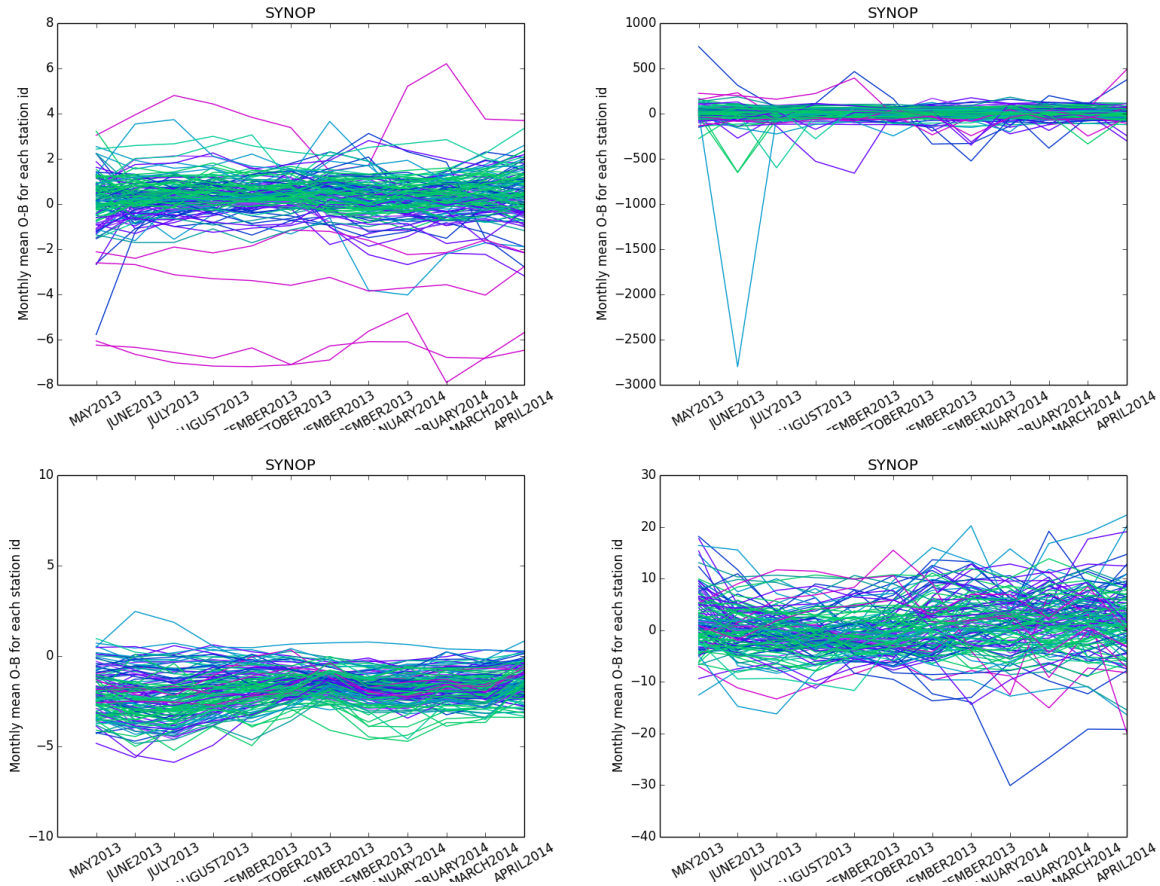

Figure 16: Monthly mean O-B for each individual SYNOP station with different colours representing different station altitude categories. 2m temperature (upper left), surface pressure (upper right), 10m wind speed (lower left) and 2m relative humidity (lower right) Green lines are low altitudes fading through to pink lines for the high altitudes.

© Crown copyright 2014 Met Office

Examples of SYNOP O and B Comparisons
2m temperature

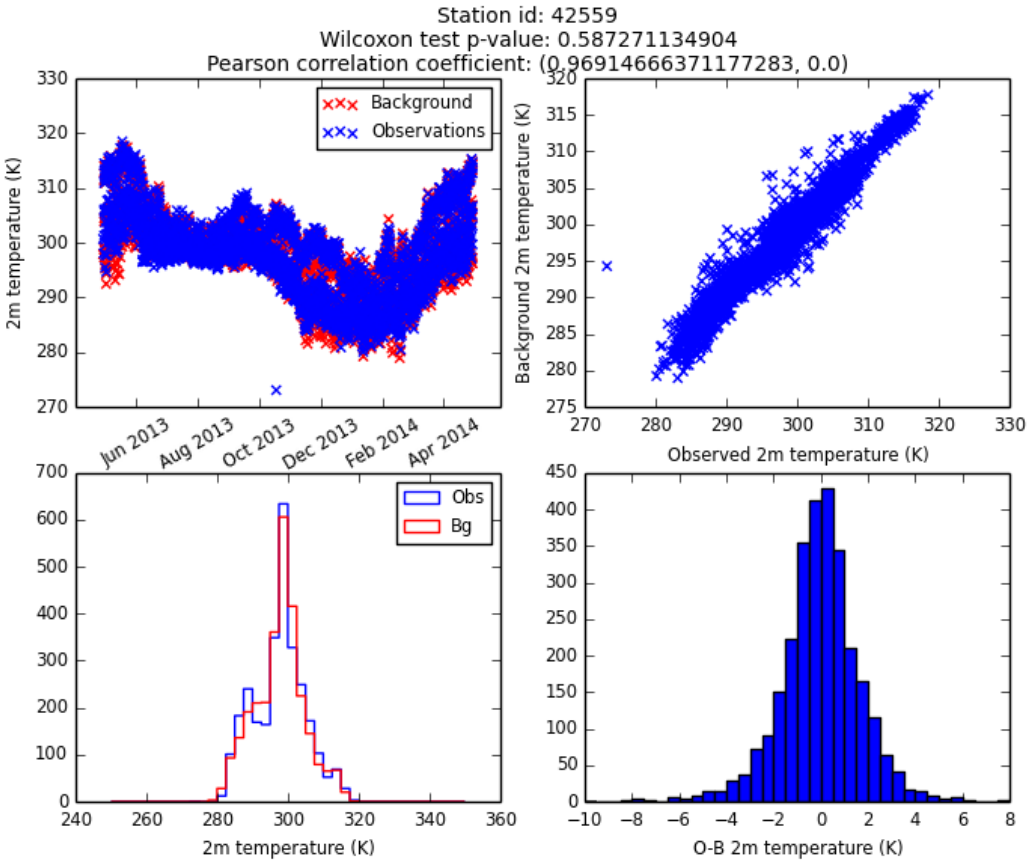

Figure 17: SYNOP station id 42559, 24.6N, 77.3E, 478m, scatter plot of observed and background 2m temperature values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

Rather than considering the region's observations as a whole, Figures 17 to 35 are examples of individual stations with SYNOP reports and compare their observed and background values and consider the distribution of O-B values for the time period May 2013 until April 2014. Figure 17 shows an example of a station which has fairly good agreement between observed and background 2m temperature, WMO station ID 42559. The high Wilcoxon signed rank test p-value suggests that the null hypothesis that the observed and background values could be from the same distribution should be retained, and the Pearson correlation coefficient close to 1 implies a strong positive relationship between the two sets of values. The distribution of O-B creates a bell-shaped curve centred about 0 as would be expected if the values are unbiased and unskewed.

Figures 18 to 20 show stations which do not have such good agreement between observed and background 2m temperatures. Station ID 43473 is an example where there is a negative bias over the year (Figure 18). The upper right plot shows that there is an approximately linear relationship between the observed and background values with moderate positive correlation. However, the lower plots show a shift of ~-5K between observed and background values.

© Crown copyright 2014 Met Office

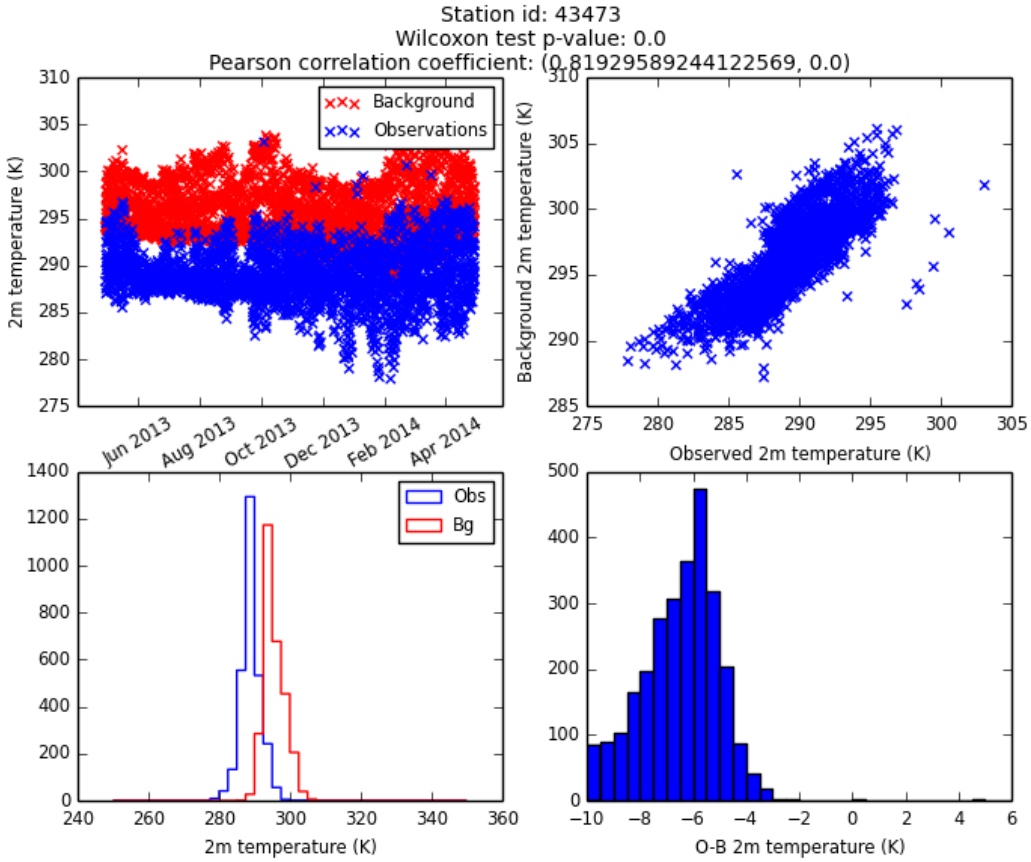

Figure 18: SYNOP station id 43473, 7.0N, 80.8E, 188m, scatter plot of observed and background 2m temperature values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

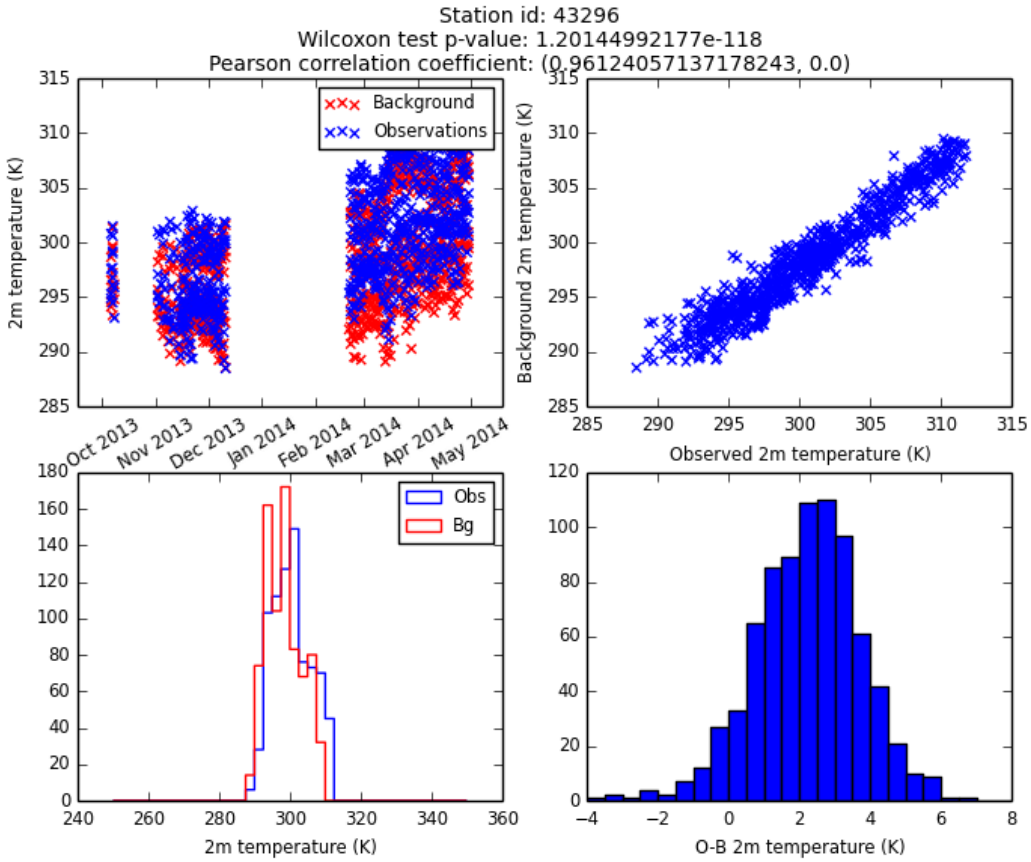

Figure 19: SYNOP station id 43296, 13.0N, 77.6E, 897m, scatter plot of observed and background 2m temperature values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

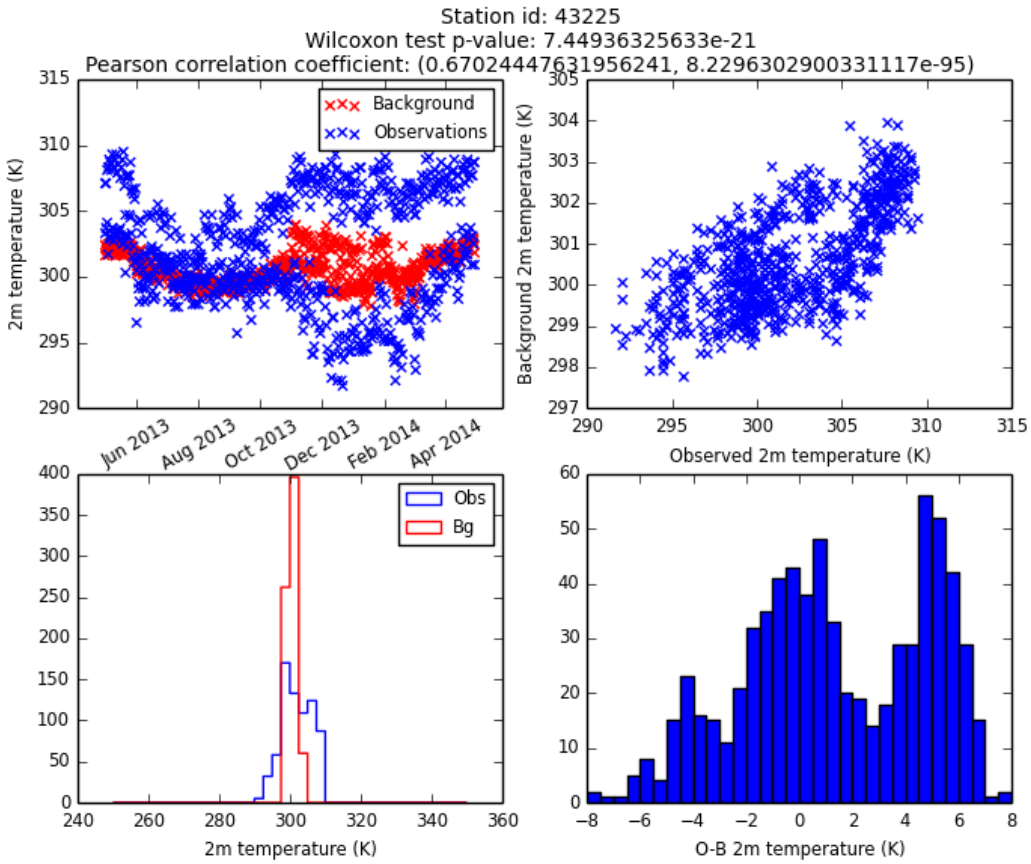

Figure 20: SYNOP station id 43225, 14.8N, 74.1E, 4m, scatter plot of observed and background 2m temperature values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

Station ID 43225 has more spread in observed 2m temperature than the background, which is most pronounced during December 2013 until March 2014 (Figure 20) when it appears that there are two bands of observation values either side of the background values. This may be due to the model maybe not having a large enough diurnal range in the background values it produces. There is a large spread in O-B values ranging from ~-6K to ~6K.

© Crown copyright 2014 Met Office

Surface pressure

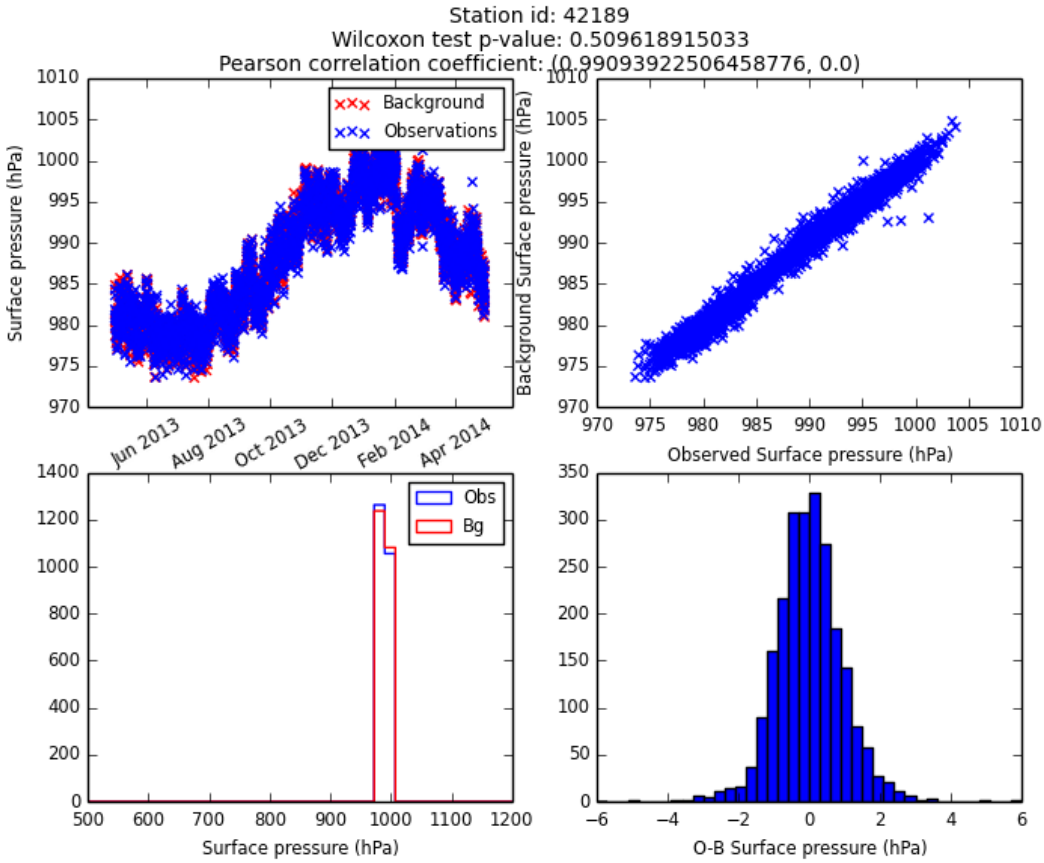

Figure 21: SYNOP station id 42189, 28.4N, 79.4E, 173m, scatter plot of observed and background surface pressure values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

Observed and background values of surface pressure at station ID 42189 have a very strong positive correlation and based on a Wilcoxon signed rank test, are likely to be from the same distribution. The distribution of O-B values is fairly symmetrical and centred at 0 (Figure 21). This is an example of a station where the observations and background agree well.

Station IDs 42498, 42701, 42921, 43087 and 43296 on the other hand, are examples of stations where the observations and background surface pressure values do not agree very well (Figures 22-26). Two of the stations have a bias through the year; ID 42498 has a negative bias and ID 42941 has a positive bias as well as the observed and background surface pressure values becoming less strongly correlated at higher pressures. Station ID 42701 has a fairly symmetrical distribution of O-B values about 0, but the observed and background values are only weakly correlated, probably because there are a number of observed values considerably outside the range of background values (Figure 23). There are months with missing data from station ID 43296, and there appears to be a change in relationship between observed and background values during the period (Figure 26).

© Crown copyright 2014 Met Office

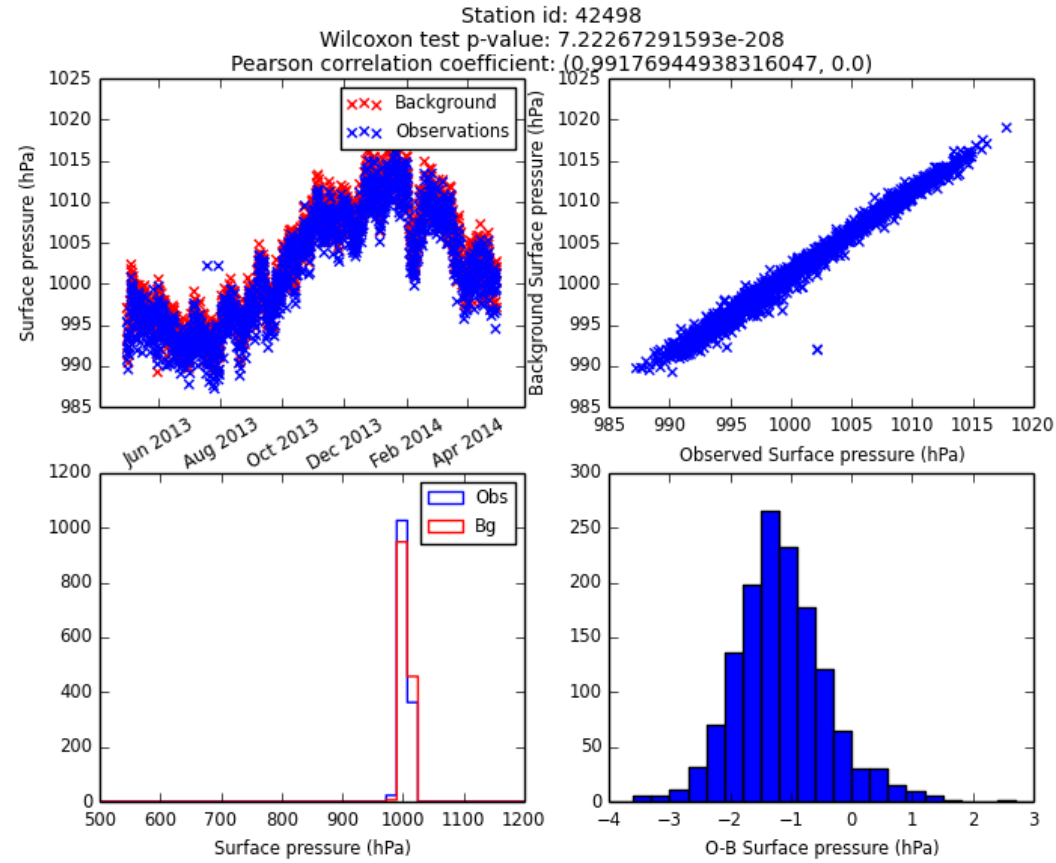

Figure 22: SYNOP station id 42498, 25.2N, 86.9E, 49m, scatter plot of observed and background surface pressure values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

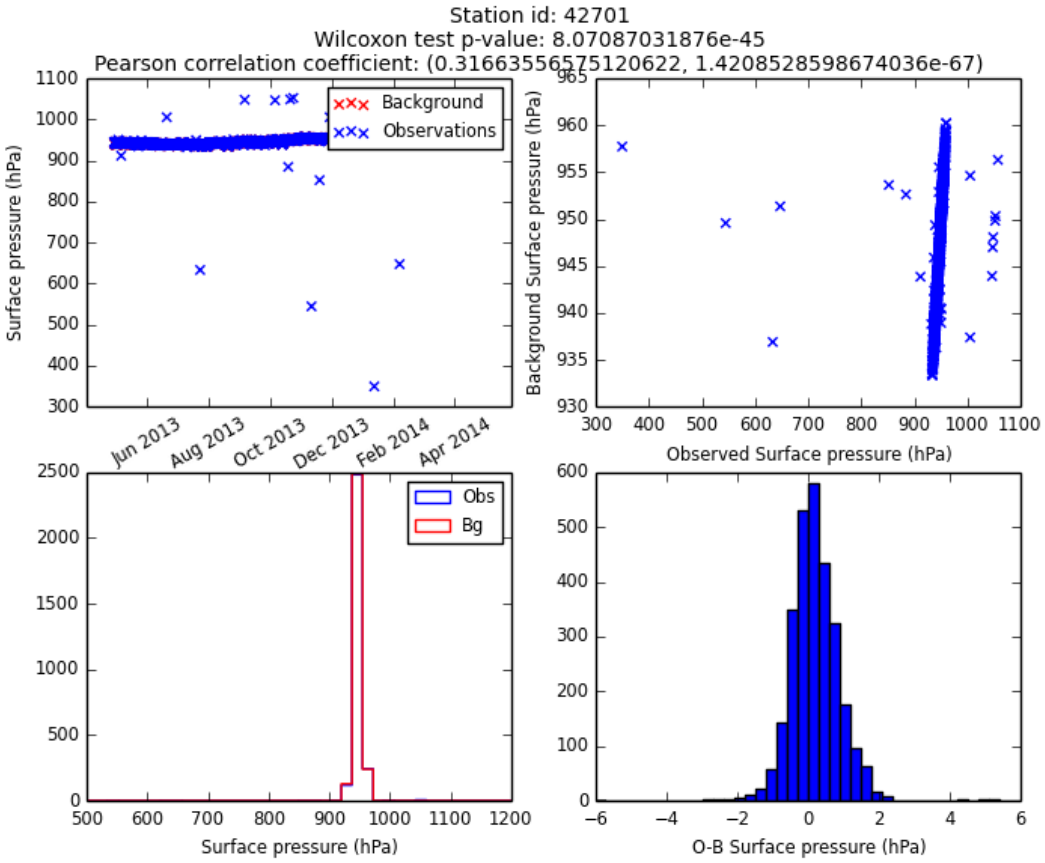

Figure 23: SYNOP station id 42701, 23.3N, 85.3E, 652m, scatter plot of observed and background surface pressure values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

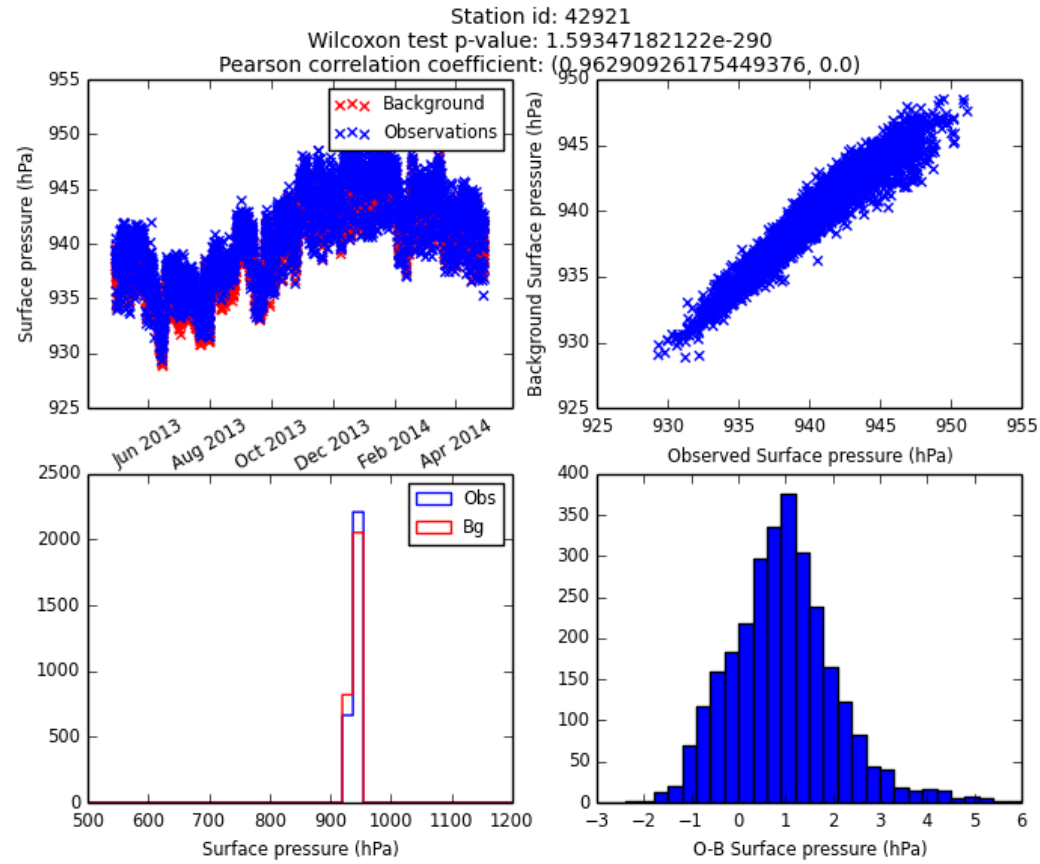

Figure 24: SYNOP station id 42921, 20.0N, 73.8E, scatter plot of observed and background surface pressure values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

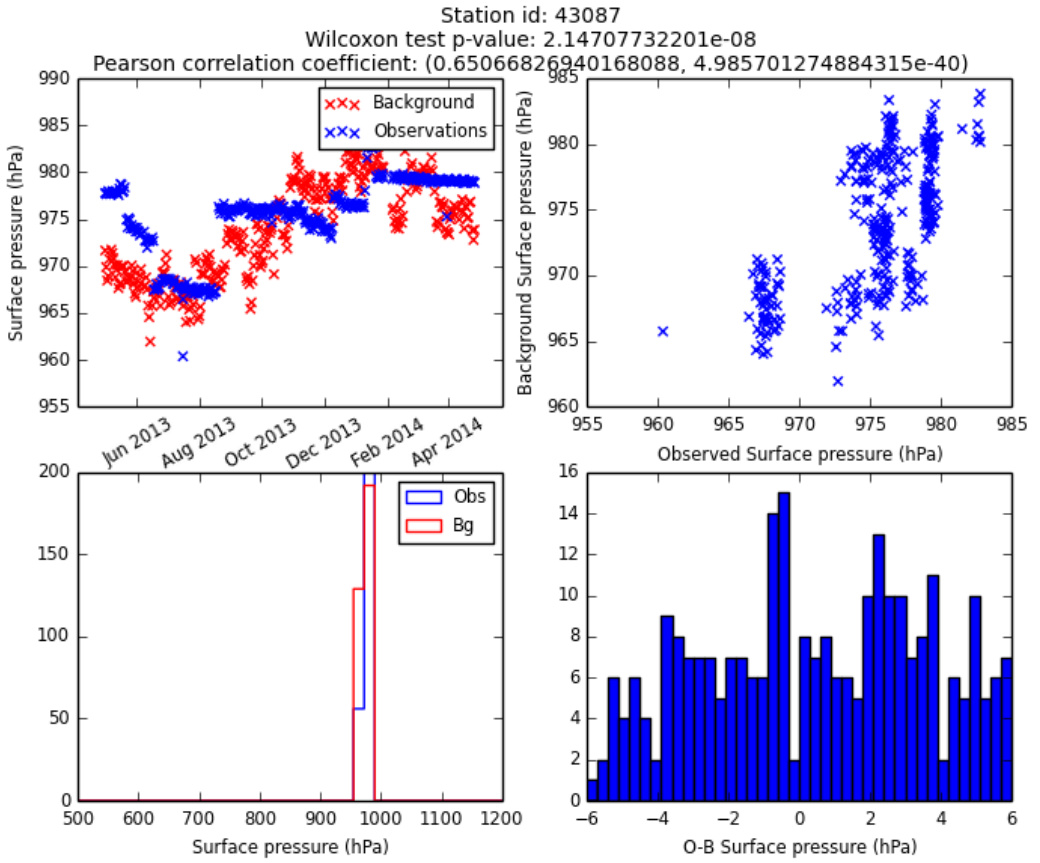

Figure 25: SYNOP station id 43087, 18.0N, 79.6E, 269m, scatter plot of observed and background surface pressure values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

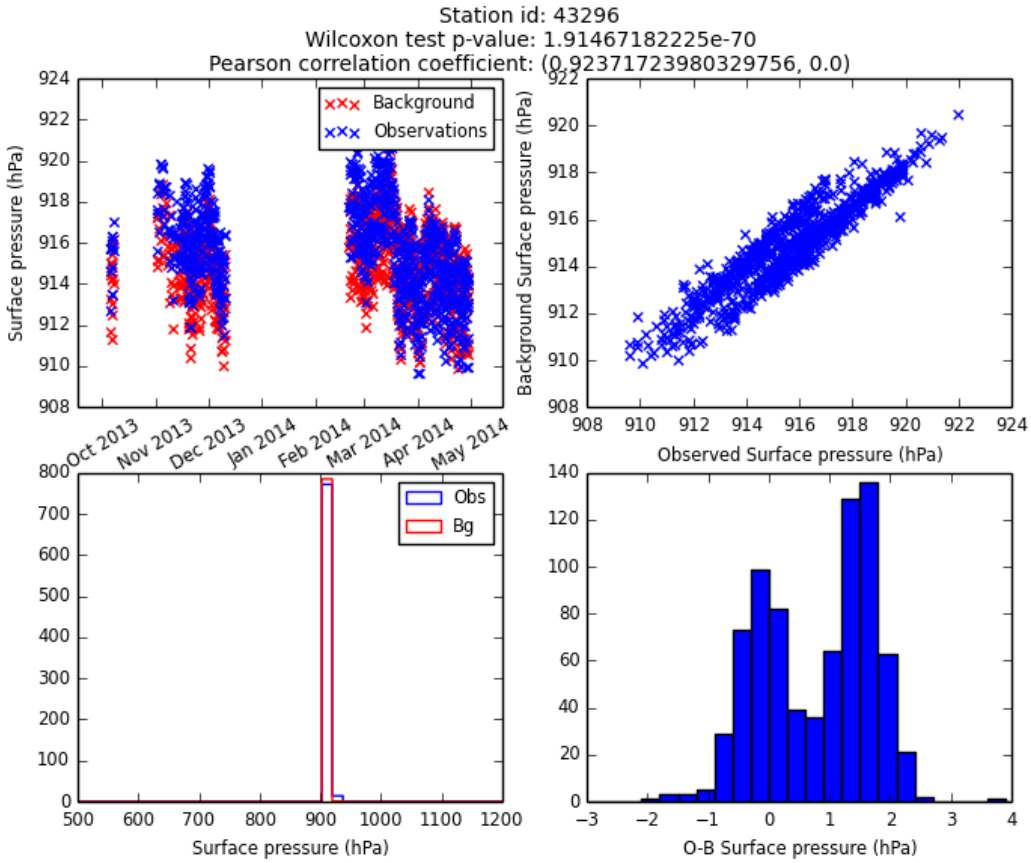

Figure 26: SYNOP station id 43296, 13.0N, 77.6E, 897m, scatter plot of observed and background surface pressure values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

From the upper left plot of Figure 26, there appears to be less bias at the end of the time period, which may be why there is bimodality in the O-B distribution. It may also be possible that the station has moved. Station ID 43087 has jumps in observed surface pressure values, which may be due to an instrument problem (Figure 25). The distribution of O-B values is not bell-shaped as would be expected if the observed and background values agreed well with each other.

© Crown copyright 2014 Met Office

Wind Speed

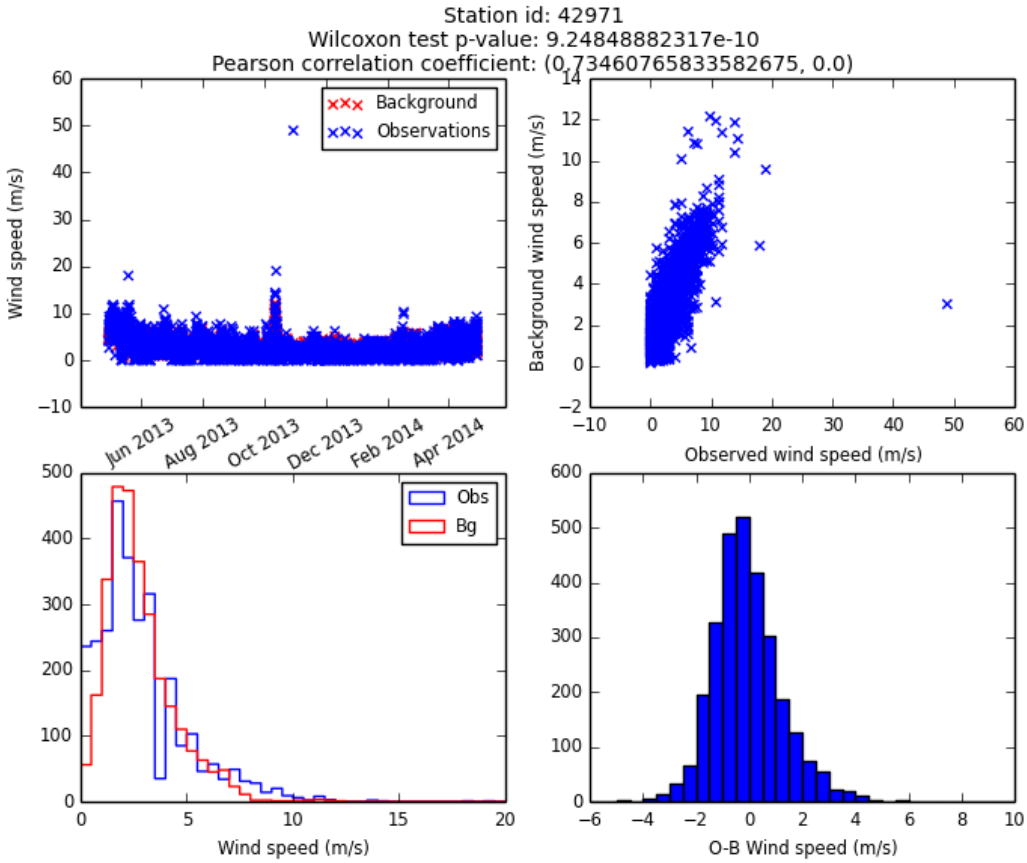

Figure 27: SYNOP station id 42971, 20.3N, 85.8E, 46m, scatter plot of observed and background 10m wind speed values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

10m wind speed observed and background values tend to disagree with each other more than the other variables considered in this report. Station ID 42971 is an example of a station which has reasonable agreement between observed and background 10m wind speeds. The histograms of each are fairly similar and the distribution of O-B is bell-shaped about 0 (Figure 27).

It has been noticed that many stations' wind speed observations are lower than the background in the Indian region (Ingleby, 2014). Station IDs 42101 and 43087 (Figures 28 and 31) are examples where this is the case. Both have a large negative bias in their O-B values; the first has many 0m/s observed values, while the latter has repeated observed 10m wind speeds of 0.5m/s throughout the year. Station ID 42398 has a negative bias in its O-B distribution (Figure 29). From the upper left hand plot, it appears that there is more difference between observed and background 10m wind speed values during November 2013 to February 2014 than the rest of the time period. Station ID 42909 also exhibits a negative bias (Figure 31). As well as this, from the upper left hand plot it seems that 3.5m/s winds are observed very infrequently. Almost all of the observed 10m wind speeds are whole or half m/s values, which may explain some of the differences between observed and background values.

© Crown copyright 2014 Met Office

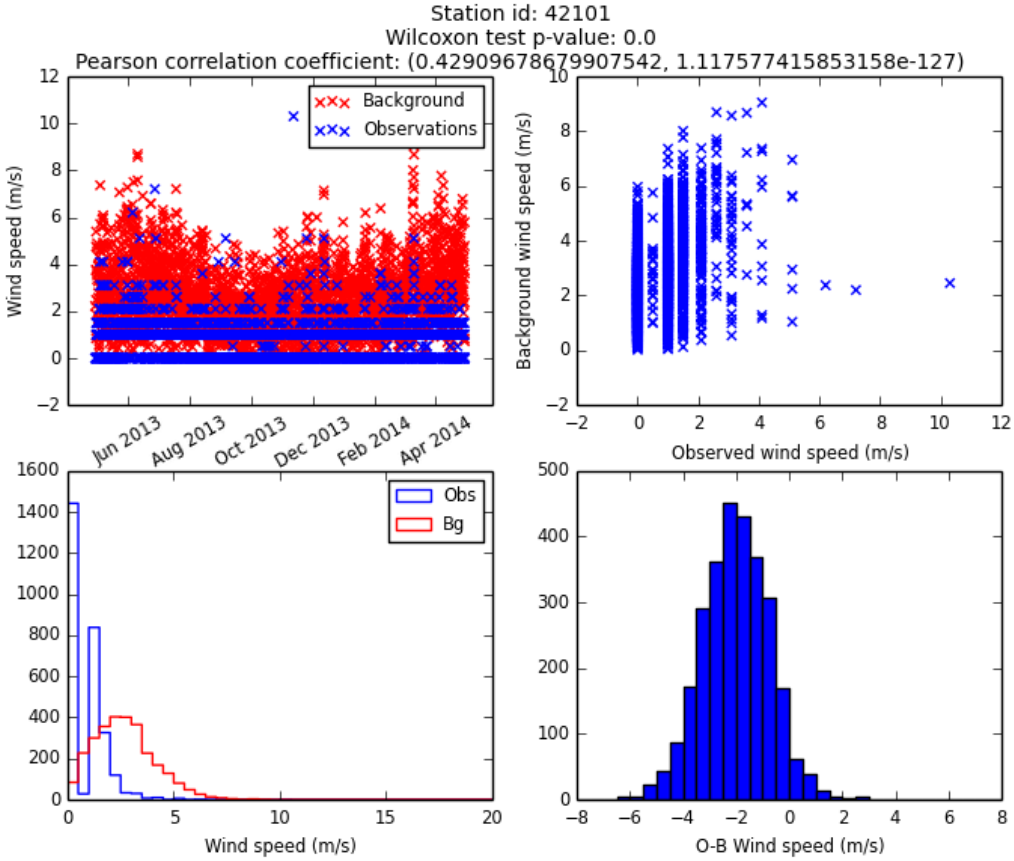

Figure 28: SYNOP station id 42101, 30.3N, 76.5E, 251m, scatter plot of observed and background 10m wind speed values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

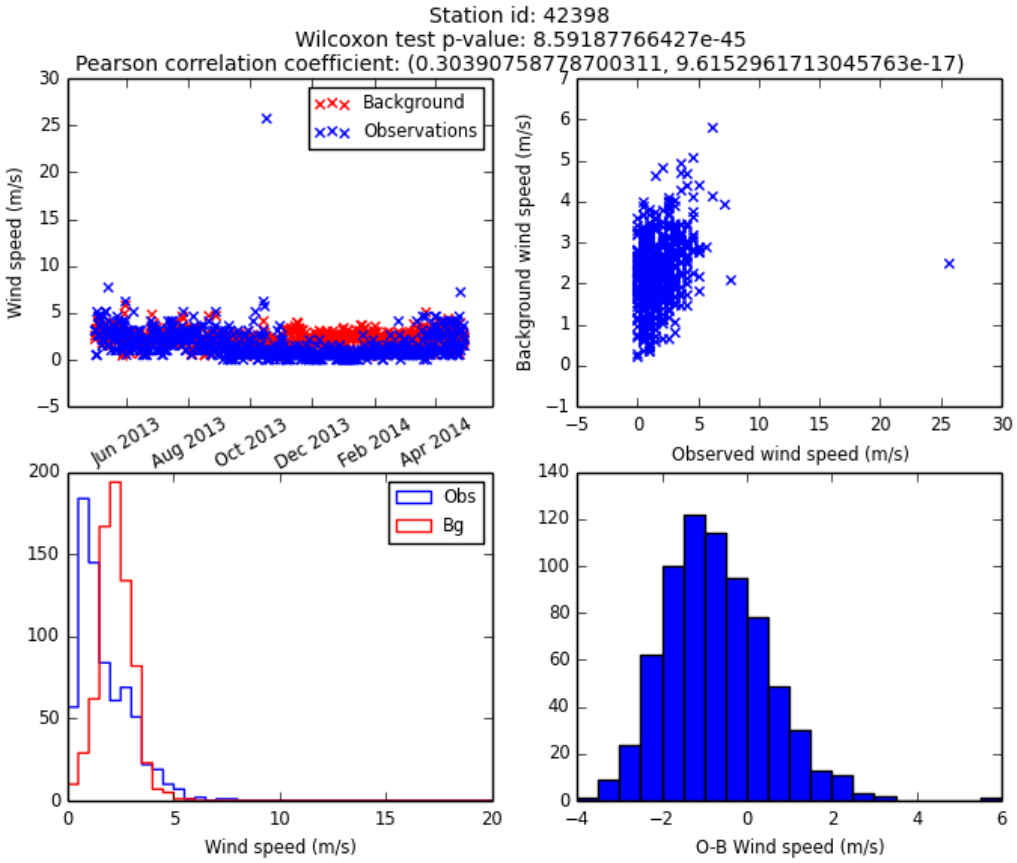

Figure 29: SYNOP station id 42398, 26.6N, 88.3E, 131m, scatter plot of observed and background 10m wind speed values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

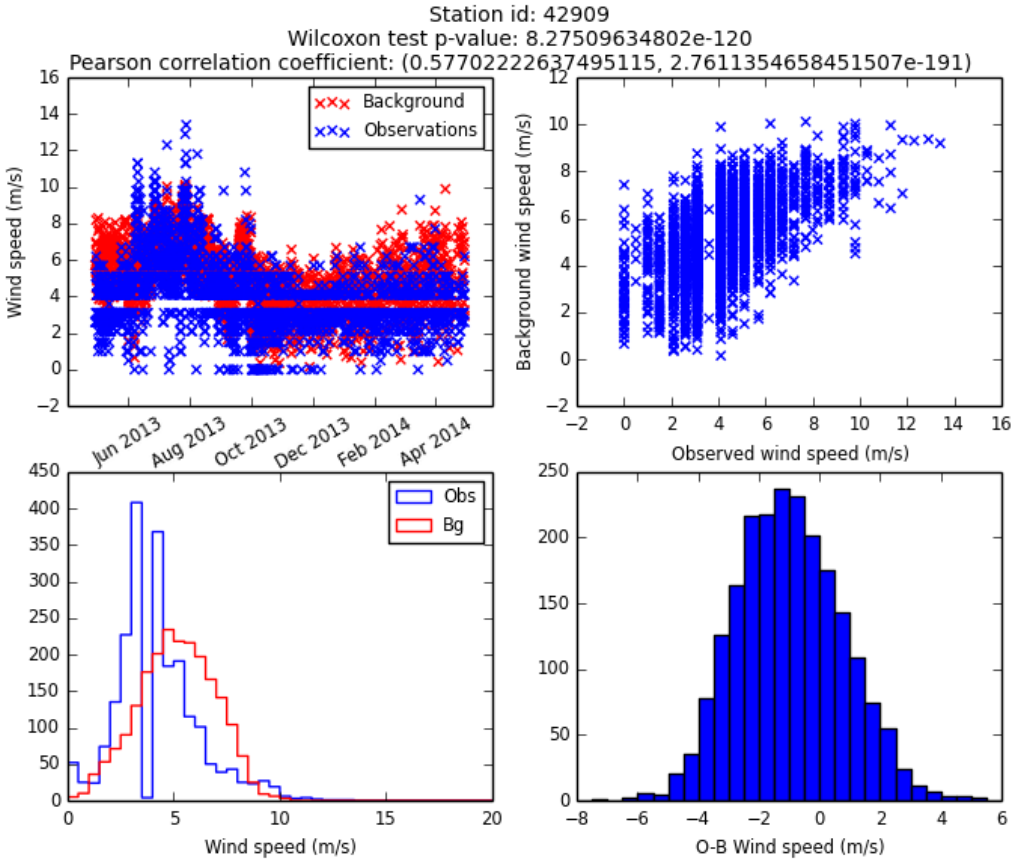

Figure 30: SYNOP station id 42909, 20.9N, 70.4E, 8m, scatter plot of observed and background 10m wind speed values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

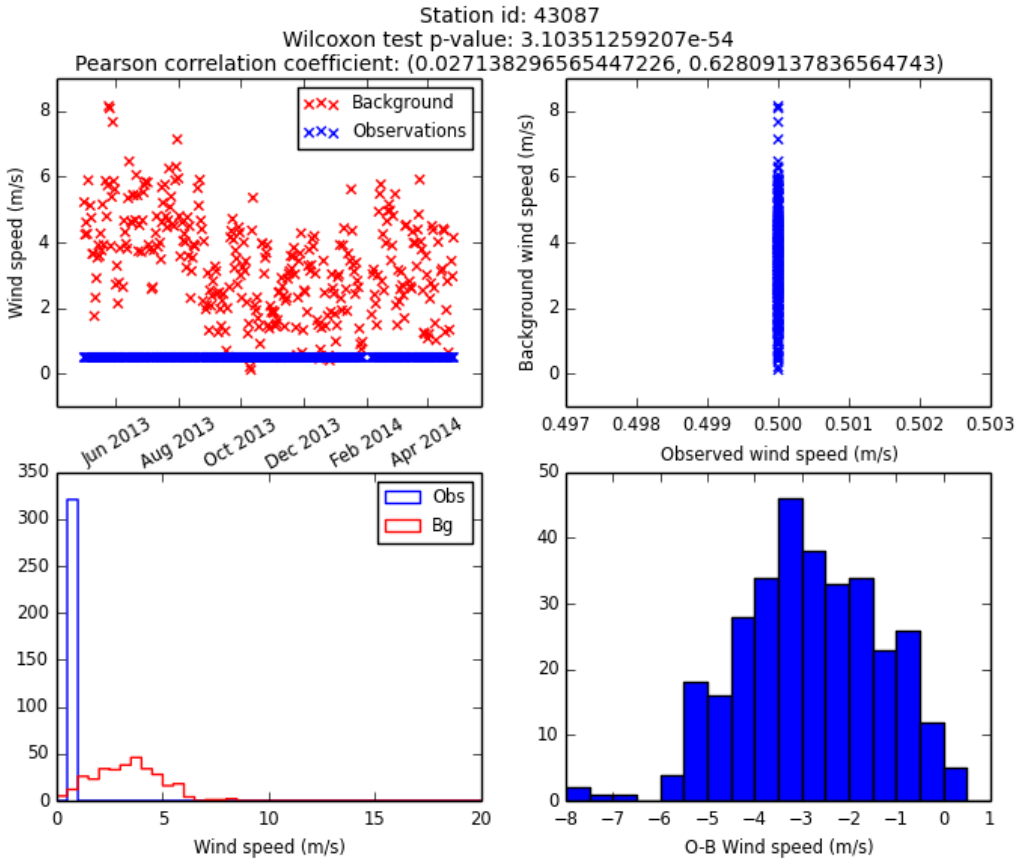

Figure 31: SYNOP station id 43087, 18.0N, 79.6E, 269m, scatter plot of observed and background 10m wind speed values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

2m Relative Humidity

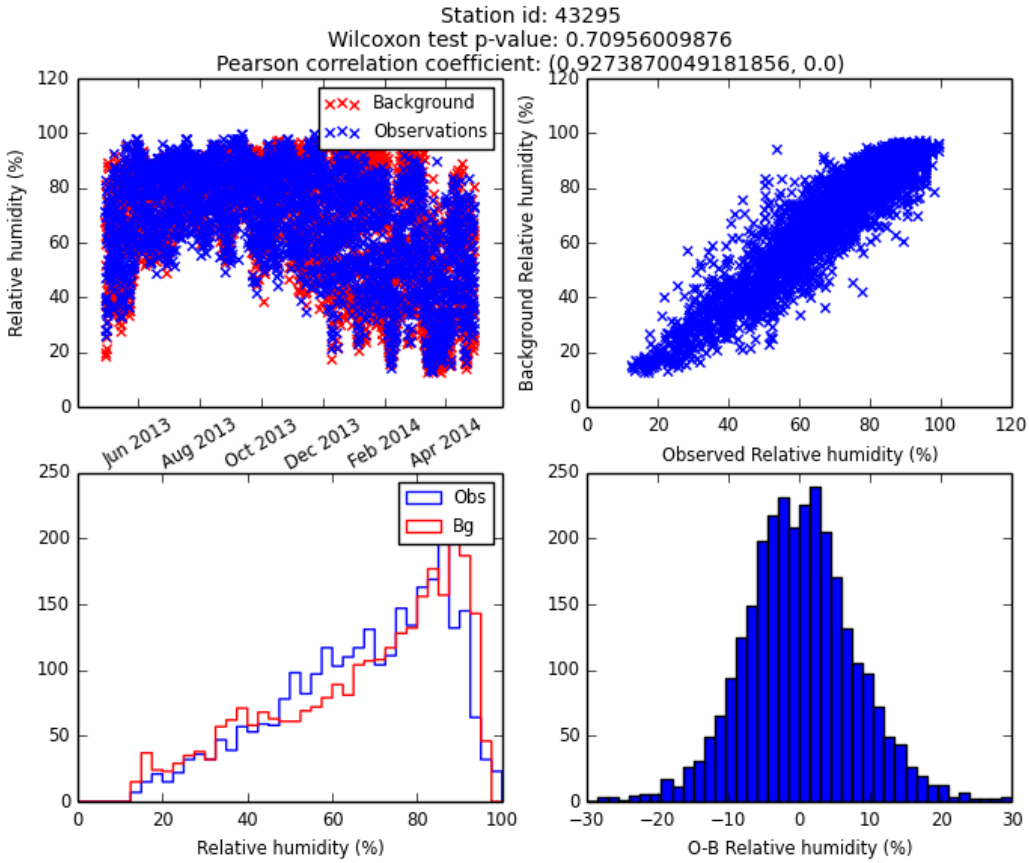

Figure 32: SYNOP station id 43295, 13.0N, 77.6E, 921m, scatter plot of observed and background 2m relative humidity values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

Observed and background 2m relative humidity values from station ID 43295 (Figure 32) have strong positive correlation and based on a Wilcoxon signed rank test the null hypothesis is retained, so they could be from the same distribution. The O-B distribution is bell-shaped and centred at 0, with high agreement between observed and background relative humidity.

Station ID 42071 has a slightly positive bias to its O-B distribution (Figure 33), and we can see that there are a large number of observations of 100% relative humidity. Conversely, station ID 42348 has a negative bias as there are more high background values and lower observed values of relative humidity (Figure 34). Despite this, there are more observed than background values of 100%. Station IDs 43157 and 43225 (Figures 35 and 36) have different spreads of observed and background 2m relative humidity values; the first has less spread in the observed values than the background, whereas the second has less spread in the background vales than the observed values. For 2m relative humidity, a number of stations seem to exhibit a pattern in the scatter plots of observed values against background values as in the upper right plot of Figure 34 for station ID 42348. The curved cloud of points implies that at lower observed values, there are more times that the background overestimated the 2m relative humidity, but at the higher values, it is underestimated.

© Crown copyright 2014 Met Office

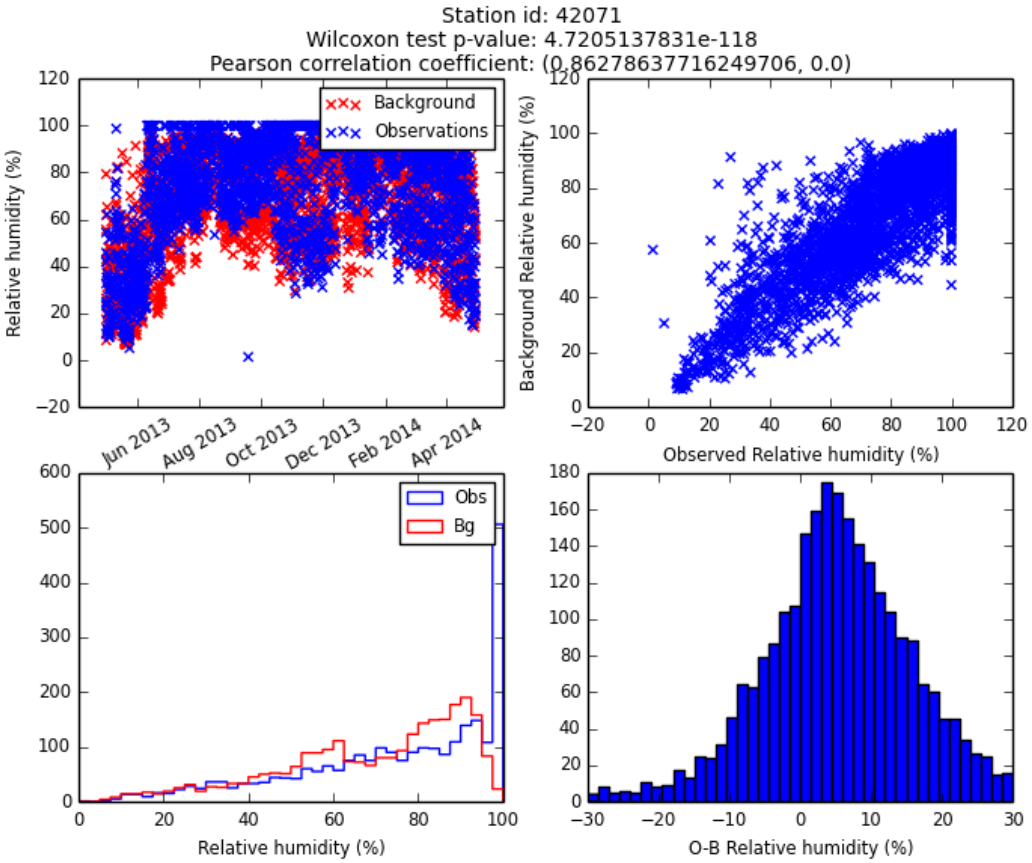

Figure 33: SYNOP station id 42071, 31.6N, 74.9E, 234m, scatter plot of observed and background 2m relative humidity values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

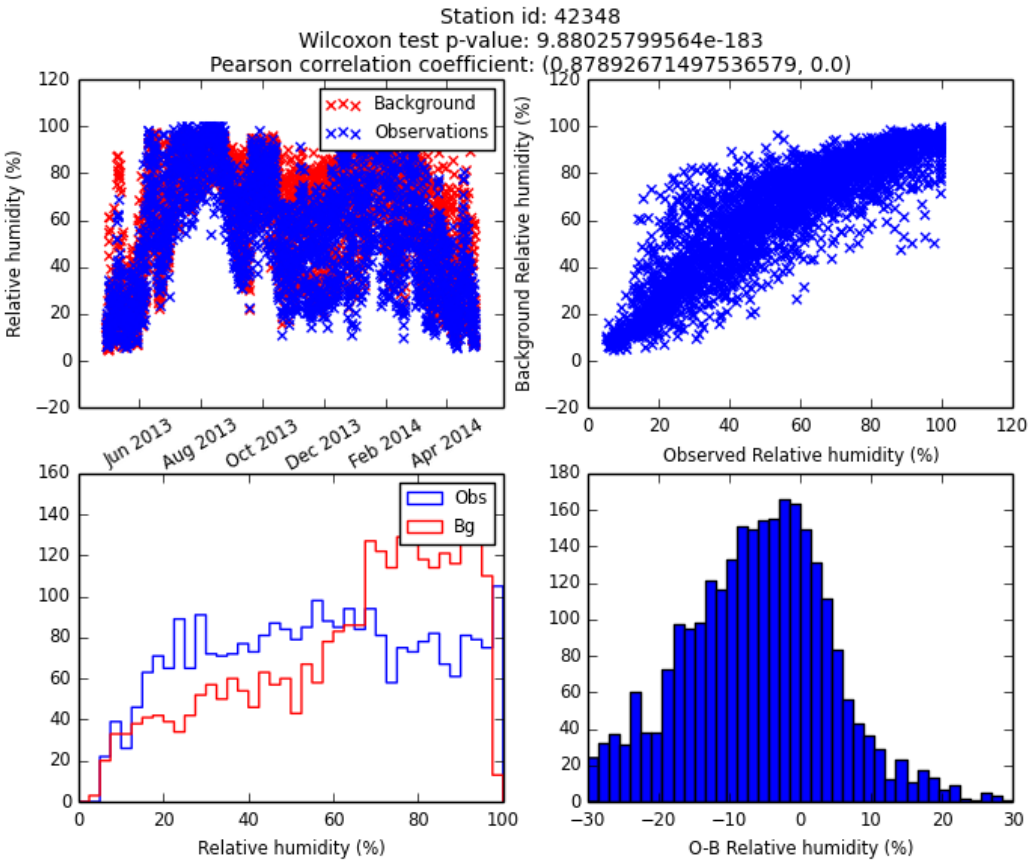

Figure 34: SYNOP station id 42348, 26.8N, 75.8E, 390m, scatter plot of observed and background 2m relative humidity values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

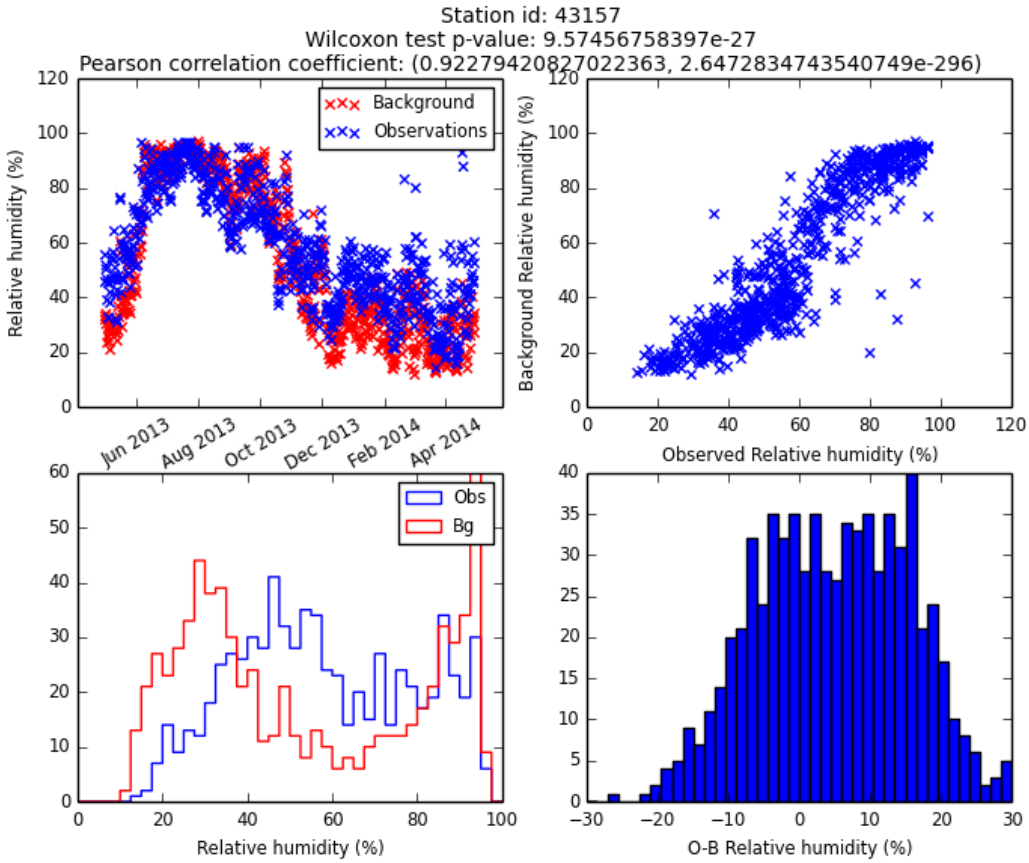

Figure 35: SYNOP station id 43157, 16.7N, 74.2E, 570m, scatter plot of observed and background 2m relative humidity values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

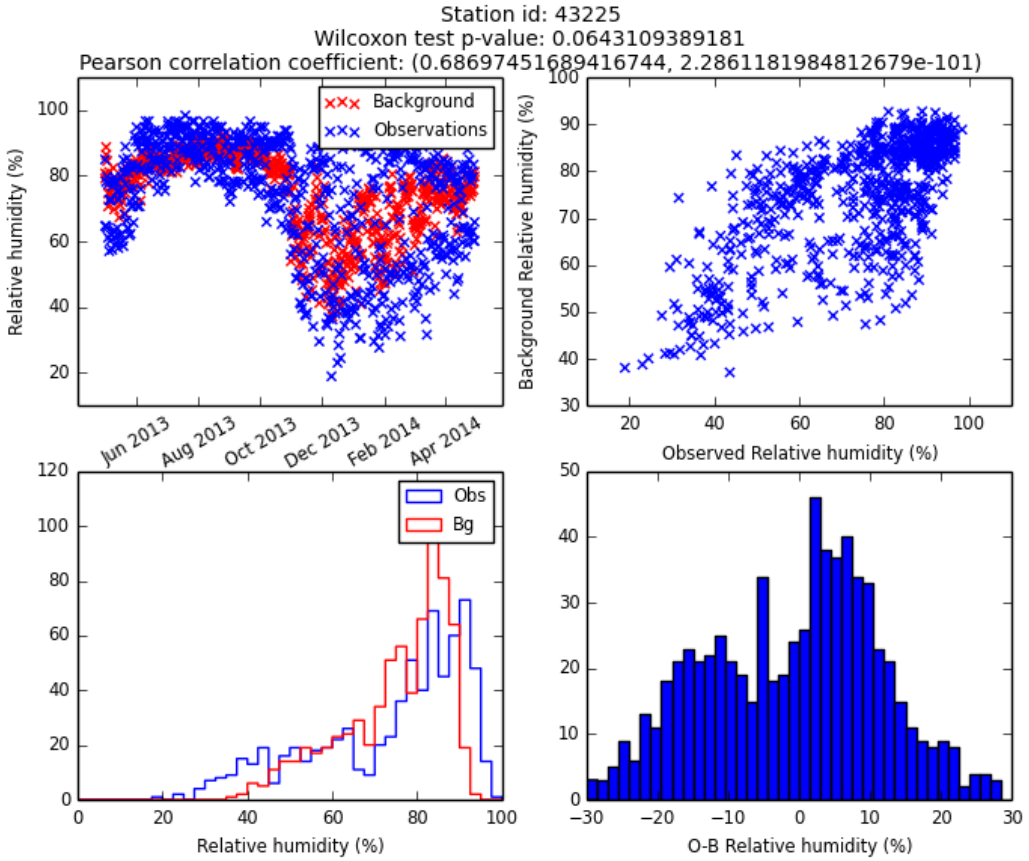

Figure 36: SYNOP station id 43225, 14.8N, 74.1E, 4m, scatter plot of observed and background 2m relative humidity values (top left), scatter plot of observed values against background values (top right), histograms of observed and background values (bottom left) and histogram of O-B values (bottom right)

© Crown copyright 2014 Met Office

TEMP O and B comparison

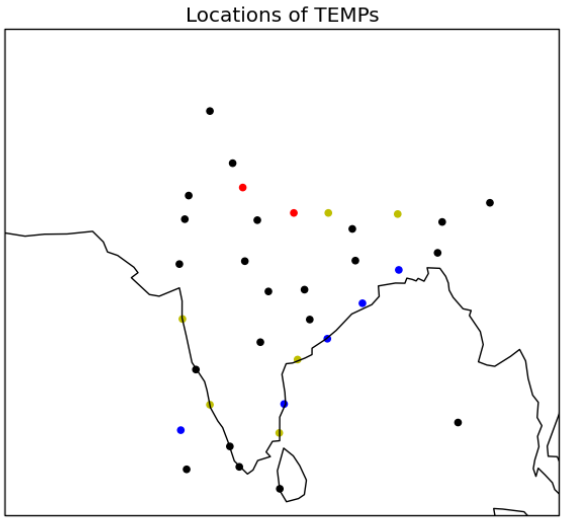

Figure 37: Locations of Indian TEMP reports. Red points show stations for which temperature, relative humidity and winds are currently permanently rejected. Yellow points show stations for which temperature and relative humidity are currently permanently rejected. Blue points show stations which have good O-B profiles in recent months. Black points are remaining stations.

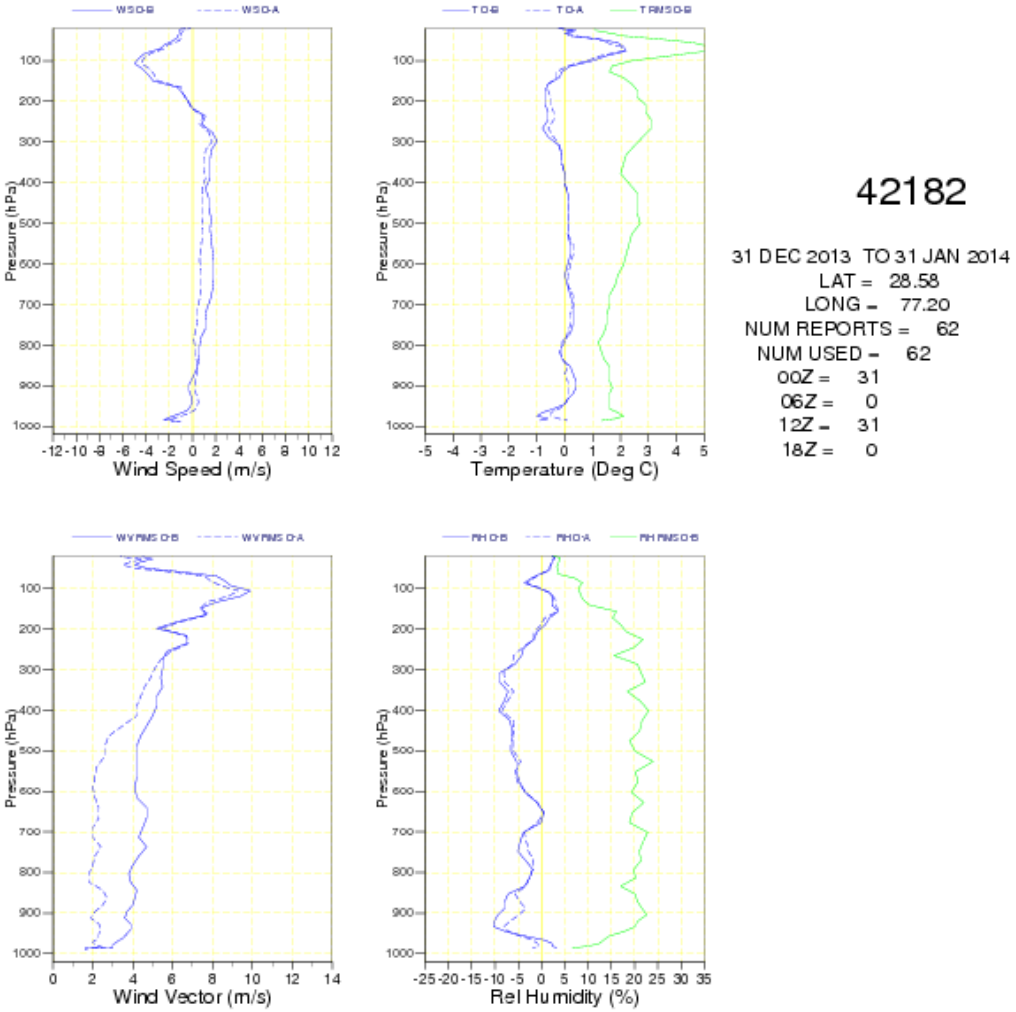

Figure 38: January 2014 O-B profiles for TEMP reports, station id 42182 for variables wind speed (top left), temperature (top right), wind vector (bottom left) and relative humidity (bottom right)

© Crown copyright 2014 Met Office

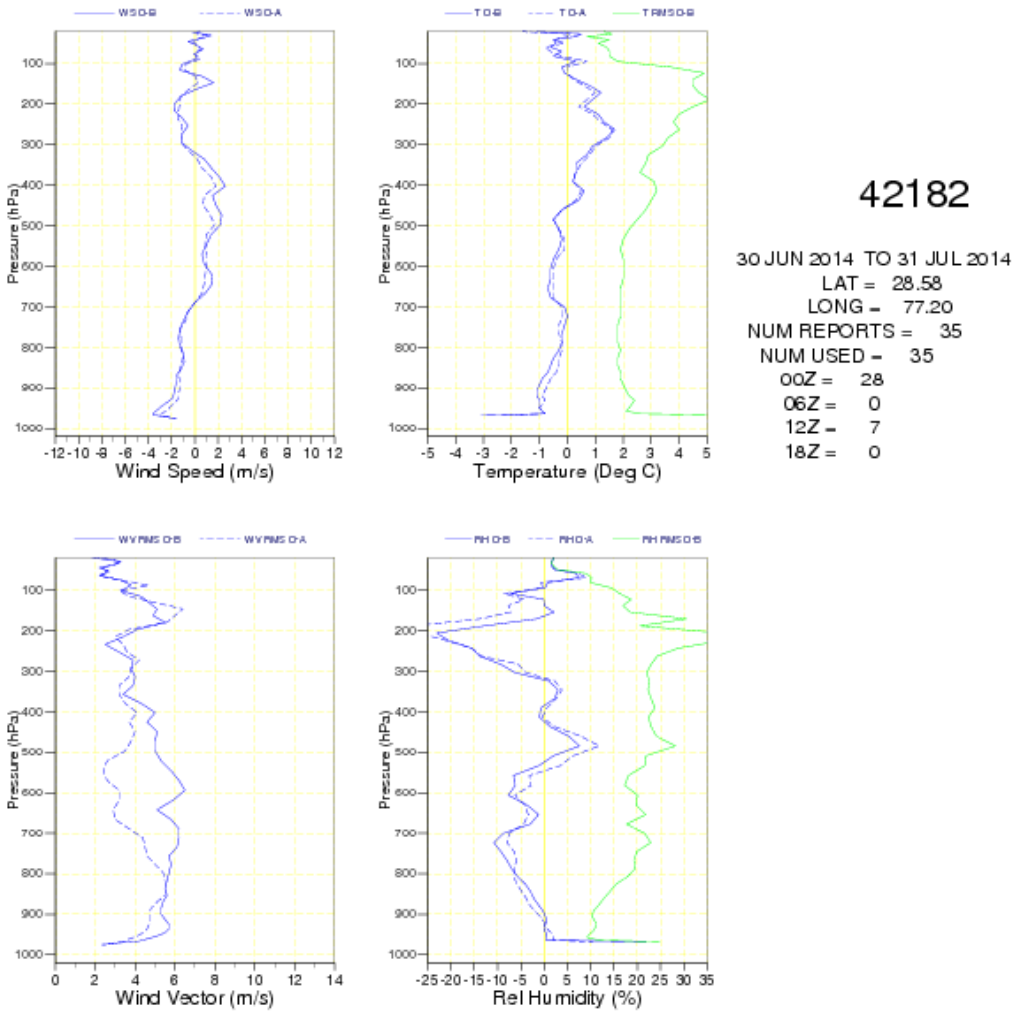

Figure 39: July 2014 O-B profiles for TEMP reports, station id 42182 for variables wind speed (top left), temperature (top right), wind vector (bottom left) and relative humidity (bottom right)

Figure 37 shows the location of Indian TEMP reports. There are eight stations whose temperature and relative humidity reports are currently on a permanent rejection list because of the quality based on their O-B statistics. The locations of these stations are shown as red and yellow points on the map. Two of these stations also have their wind reports rejected, and are those shown in red. There are a few stations, shown in blue, which in recent months have been suggested as examples of good quality based on their O-B profiles (C. Parrett, personal communication).

Temperature, relative humidity and wind reports from station ID 42182 are rejected. O-B profiles for the months of January 2014 and July 2014 are shown in Figures 38 and 39 respectively. There are large RMS O-B values for temperature and relative humidity during both months, particularly higher in the atmosphere. There is a negative bias for wind speed near the surface, but higher up there is a positive bias; between 900hPa and 200hPa for January 2014 and between 700hPa and 300hPa for July 2014. Relative humidity has a negative bias through most pressure levels, which is very large during July 2014 at ~ 200hPa.

© Crown copyright 2014 Met Office

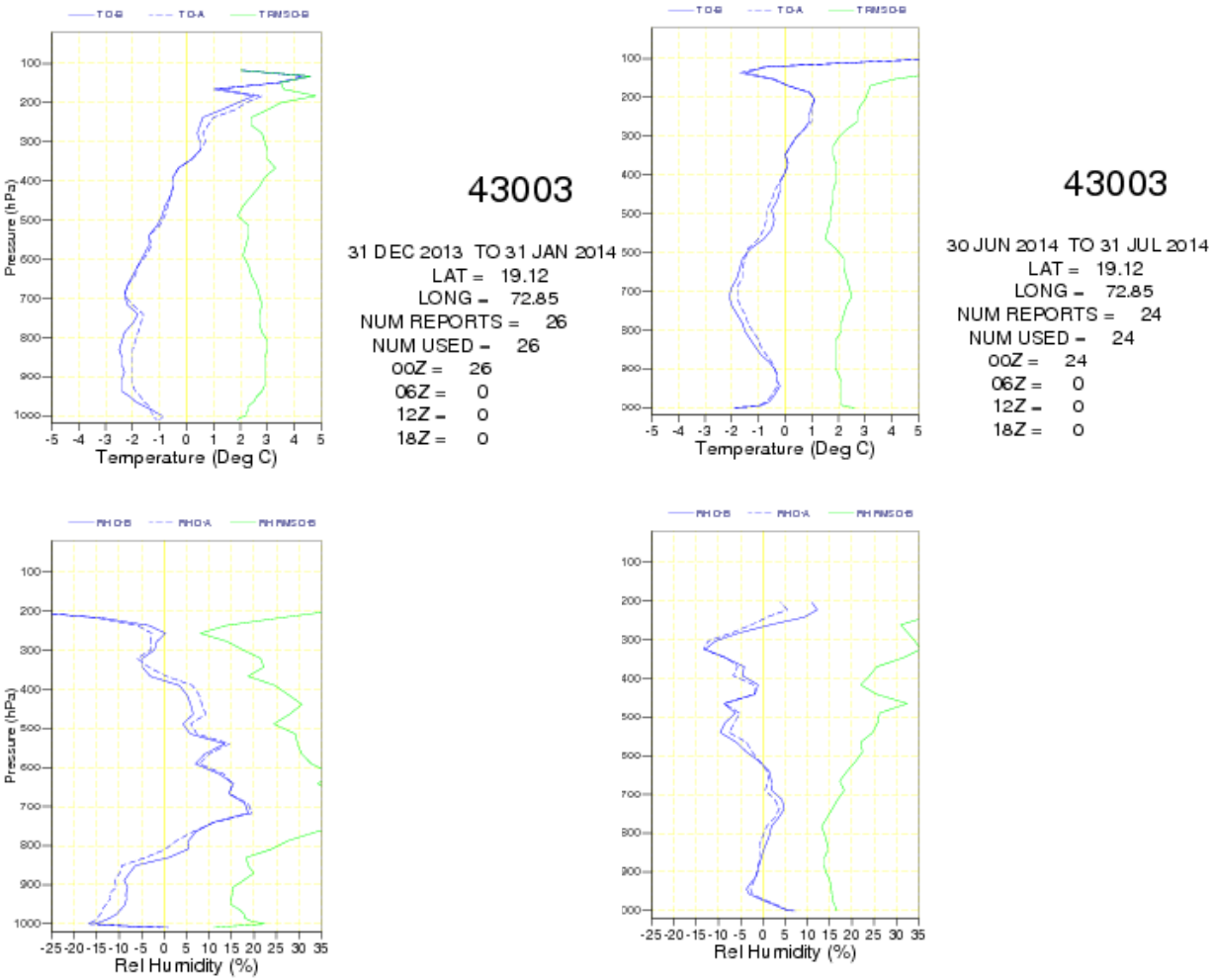

Figure 40: January (left column) and July (right column) 2014 O-B profiles for TEMP reports, station id 43003 for temperature (top row) and relative humidity (bottom row)

Temperature and relative humidity reports from station ID 43003 are currently rejected due to their quality. Figure 40 shows O-B profiles for both variables for January 2014 (left column) and July 2014 (right column). In both months there is a negative bias for temperature up to approximately 300hPa, and also a large RMS of O-B (~2°C) which is slightly larger in January 2014 than July 2014. There is also a large RMS of relative humidity O-B, exceeding 35% at some heights in January 2014. The bias profile is dissimilar for the months, with a negative bias at lower (up to ~800hPa) and higher (above ~400hPa) altitudes and a positive bias at levels between those for January, but little bias in July to approximately 600hPa and a negative bias higher up.

© Crown copyright 2014 Met Office

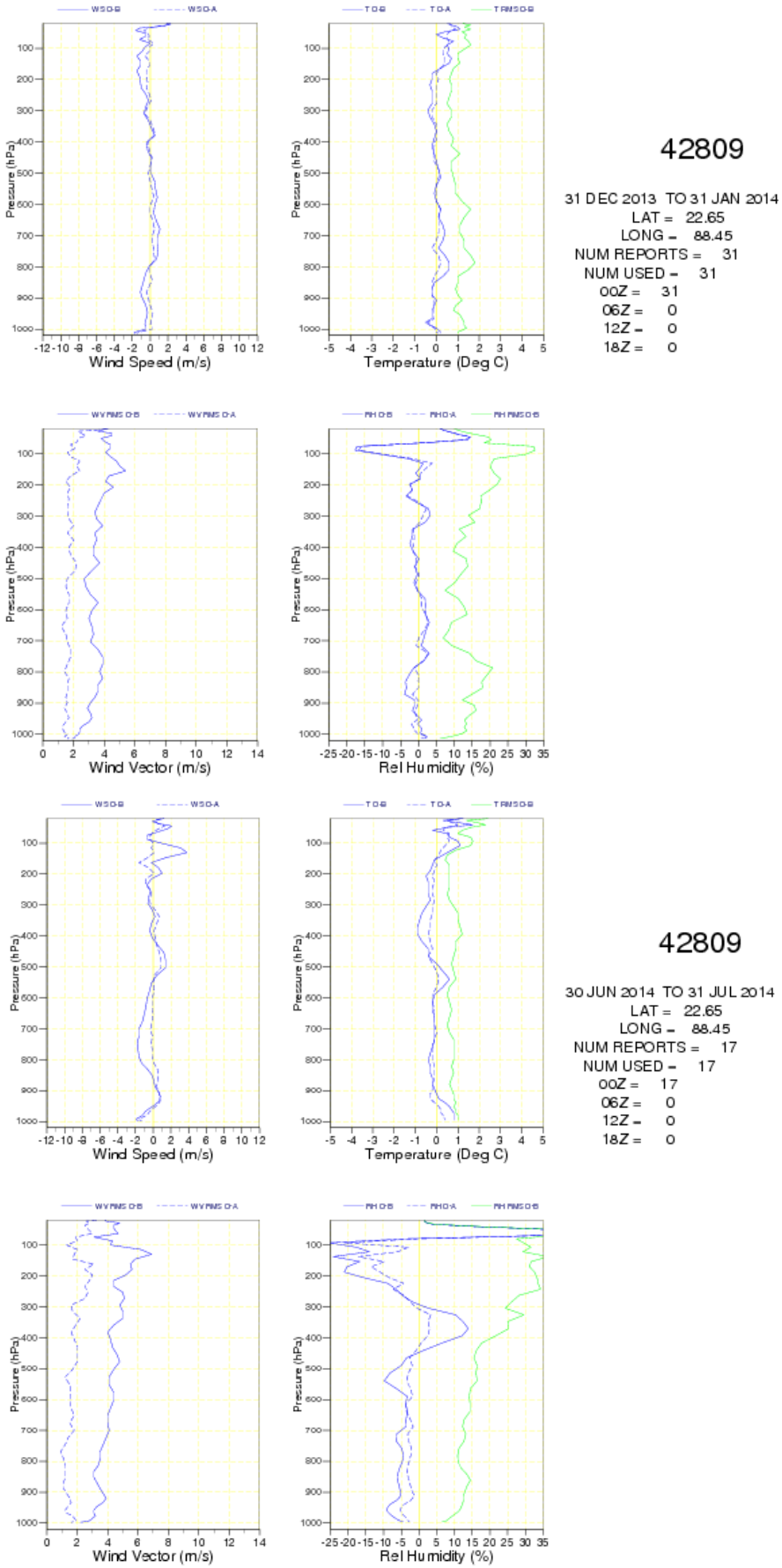

Figure 41: January (top set of 4 plots) and July (bottom set of 4 plots) 2014 O-B profiles for TEMP reports, station id 42809 for variables wind speed (top left), temperature (top right), wind vector (bottom left) and relative humidity (bottom right)

© Crown copyright 2014 Met Office

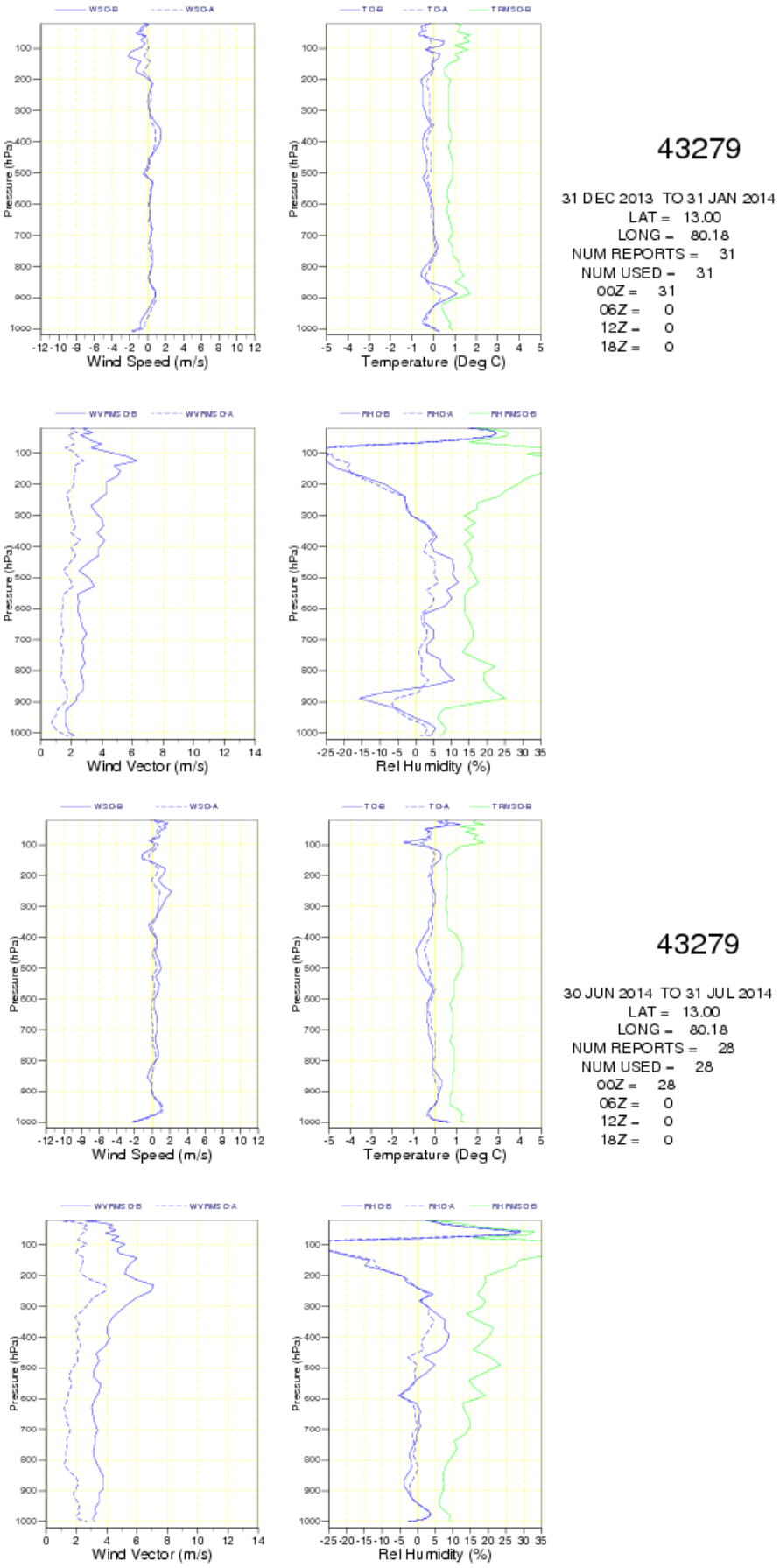

Figure 42: January (top set of 4 plots) and July (bottom set of 4 plots) 2014 O-B profiles for TEMP reports, station id 43279 for variables wind speed (top left), temperature (top right), wind vector (bottom left) and relative humidity (bottom right)

© Crown copyright 2014 Met Office

Reports from station IDs 42809 and 43279, however, are assimilated subject to the usual blacklisting and rejection criteria. For both stations, during January and July 2014, there is low bias for temperature and wind speed. The bias for relative humidity is also generally smaller than for the previous stations mentioned whose reports are not assimilated (Figures 41 and 42). The RMS O-B values for temperature are about half of those for station IDs 42182 and 43003, and are fairly constant at about 1°C as pressure reduces with height. Relative humidity RMS O-B values are also slightly smaller for station IDs 42809 and 43279 (~15% compared to ~20%).

© Crown copyright 2014 Met Office

CONCLUSIONS

In summary, surface data for 2m temperature, 2m relative humidity and surface pressure from SYNOP, METAR, ship and buoy reports are assimilated if they do not meet rejection criteria, but mobile SYNOP reports from the region 65E to 90E and 5N to 35N are not. Reports of winds from the Tropics, 30S to 30N, are rejected. There are 35 stations in the region with TEMP reports, 8 of which have temperature and relative humidity rejected, and 2 which have winds rejected.

Comparing surface data with model background for the region as a whole, 2m temperature observed and background values tend to be well correlated, as do surface pressure observed and background values. However, using Wilcoxon signed rank tests to determine if they are from the same distribution, none of 2m temperature, 2m relative humidity, 10m wind speed or surface pressure for the region as a whole are likely that their observed and background values are from the same distribution as each other. In all cases, the null hypothesis that they are from the same distribution was rejected in favour of the alternative hypothesis.

There is a positive bias for SYNOP reports of 2m temperature, with the highest magnitude differences in the northern part of the region, and there is less bias for ship and buoy reports. Surface pressure reports have unbiased O-B, but with some high magnitude values in mountainous areas. There is a slight negative bias for SYNOP reports of 2m relative humidity, although buoys have a positive bias. There is reduced spread in the O-B values during the summer, but the distribution is roughly symmetrical throughout the year May 2013 to April 2014. There are some differences with season as during the winter there are more stations with positive bias and during summer more with negative bias. The observed and background relative humidity are moderately positively correlated.

Observed and background wind speeds are positively correlated, but more weakly than the other three variables and there is a large negative bias of ~-1.5m/s from SYNOP reports, with ~80% of the O-B values being negative. Buoy and ship reports have smaller bias than SYNOP reports, with ships having a slight positive bias, and the mean O-B for buoys being close to 0m/s. The majority of stations have a negative bias, but there are a few coastal stations exhibiting positive bias. There are many very low magnitude observed wind speeds.

Most stations have strong positive correlations between observed and background 2m temperature and 2m relative humidity values, although island stations tend to be more weakly correlated. The majority of stations have strong positive correlation between observed and background surface pressure, which is weaker in mountainous areas. The correlations between observed and background values of wind speed at SYNOP stations tend to be weaker, with approximately half of the stations in the region having moderate positive correlation, although uncorrelated at some mountainous stations and strong positive correlations at most stations within

© Crown copyright 2014 Met Office

the Indian peninsula. There are very few stations for any of the variables where the observed and background values are likely to be from the same distribution. This is considering May 2013 until April 2014 as a whole though and there may be months or seasons within this time period for which there is more similarity between observed and background values.

For 2m temperature reports the stations which have the most negative bias tend to be from the lower latitudes, and those with the largest magnitude bias tend to be at high altitudes. Most of the stations with negative bias for wind speed during summer months tend to be at low latitudes and those with most negative bias during winter seem to be at low altitudes.

In Section 3, there were examples of stations which had good and poor agreement between observed and background values. Those with less agreement were poor in different ways, for example, bias, spread, asymmetry of O-B distribution, missing data for time periods, strings of repeated or similar values, many low values (for wind speed) and reported observed values at set increments (e.g. at half or whole m/s only). These could be due to problems with either the observations or the model background. Possible explanations could be model bias, instruments stuck at a certain value, mast heights being at different heights to those reported, model topography not matching reality due to resolution, local effects not captured by the model, observations being reported in alternative units or the model not capturing extremes which are observed.

Section 4 contained some examples of stations with TEMP reports which were unbiased with small RMS of O-B values and also some which were biased with large RMS of O-B values. There were some differences between January and July 2014 O-B profiles, and relative humidity typically had a large RMS of O-B. Some of the better stations had RMS of temperature O-B values of half the magnitude which stations with poor agreement between observations and background had.

The next step could be to attempt to determine where the large O-B values come from, i.e. whether there are problems with the observations or the model background; it is likely to be a combination and therefore difficult to separate them. It could be relevant to consider how O-B changes with varying weather conditions, for example during monsoon onset the behaviour of O-B may differ compared with how they behave post-monsoon. The model may be better at representing particular conditions than others, and observations may be more accurate within certain ranges of values. Using a longer time period may help with this. It could be useful to know of a few stations where the mast heights are known to be at the correct height, with regular frequent observations reported so that they can be used as a comparison for other stations with how they compare to the model background and each other. Another possibility could be to compare with another region of the Tropics.

© Crown copyright 2014 Met Office

REFERENCES

Fouilloux, A., 2010, [ONLINE] Available at http://nwmstest.ecmwf.int/newsevents/meetings/workshops/2010/high_performance_computing_14th/presentations/Fouilloux.pdf, Accessed 2 January 2015

Ingleby, B. (2014), Global assimilation of air temperature, humidity, wind and pressure from surface stations. Q.J.R. Meteorol. Soc.. doi: 10.1002/qj.2372

Scipy.org, 2014, [ONLINE] Available at: http://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.wilcoxon.html, Accessed 2 January 2015.

Rajagopal E N, Iyengar G R, George J P, Das Gupta M, Mohandas S, Siddharth R, Gupta A, Chourasia A, Prasad VS, Aditi, Sharma S and Ashish A. 2012. 'Implementation of Unified Model based Analysis-Forecast System at NCMRWF.' Technical Report NMRF/TR/2/2012. NCMRWF, INDIA.

Wilks DS. 2006. Statistical Methods in the Atmospheric Sciences. 2d ed. International Geophysics Series, Vol. 91, Academic Press, 35 pp

© Crown copyright 2014 Met Office

© Crown copyright 2014 Met Office

Met Office
Fitzroy Road, Exeter
Devon EX1 3PB
United Kingdom

Tel (UK): 0370 900 0100 (Int): +44 1392 885680
Fax (UK): 0370 900 5050 (Int): +44 1392 885681
enquiries@metoffice.gov.uk
www.metoffice.gov.uk